

**IL DIBATTITO SULLA PLAUSIBILITA'
DELL'ESISTENZA DI "MIGLIORI CONTINUANTI" PER
L'IDENTITA' PERSONALE.**

INDICE.

INTRODUZIONE.....	4
1 – CENNI STORICI.....	8
1.1 – Locke.....	8
1.2 – Leibniz.....	12
1.3 – Joseph Butler e Thomas Reid.....	15
1.4 – Conclusioni.....	19
2 – LA RICERCA DI CRITERI PER L’IDENTITA’ PERSONALE.....	20
2.1 – Strawson e la persona intesa come “Basic Particular”	21
2.2 – L’approccio fisicalista al problema dell’identità personale.....	27
2.3 – La “standard view” e i criteri di natura psicologica.....	31
2.4 – Conclusioni.....	37
3 – IL CRITERIALISMO PSICOLOGICO ED IL CONCETTO DI IDENTITA’.....	38
3.1 – La dipendenza sortale dell’individuazione e la conseguente assolutezza e determinatezza del concetto di identità.....	39
3.2 – L’identità degli artefatti e l’only x and y principle.....	45
3.3 – Le Teorie del miglior candidato e l’identità personale.....	49
3.4 – Il tentativo di refutazione dell’only x and y principle da parte di Nozick.....	55
3.5 – Gareth Evans e l’obiezione di vaghezza.....	62
3.6 – Conclusioni.....	67

4 – LA PROPOSTA DI NOONAN E LE SUE DIFFICOLTA’.....	69
4.1 – In difesa del Reduplication Argument.....	69
4.2 – La teoria di Noonan.....	74
4.3 – Considerazioni finali sul non criterialismo.....	78
4.4 – Conclusioni.....	82
CONCLUSIONI.....	83
BIBLIOGRAFIA.....	86

INTRODUZIONE

“Metaphysics has been often revisionary, and less often descriptive. Descriptive metaphysics is content to describe the actual structure of our thought about the world, revisionary metaphysics is concerned to produce a better structure. The productions of revisionary metaphysics remain permanently interesting, and not only as key episodes in the history of thought. Because of their articulation, and the intensity of their partial vision, the best of them are both intrinsically admirable and of enduring philosophical utility. But this last merit can be ascribed to them only because there is another kind of metaphysics which needs no justification at all beyond that of inquiry in general. Revisionary metaphysics is at the service of descriptive metaphysics.”
(P.F. Strawson. 1959. Pag.9, 1-12)

Nelle primissime righe del suo libro *Individuals. An essay in descriptive metaphysics*. (1959) P.F. Strawson traccia una linea di demarcazione tra quella che chiama una metafisica revisionista ed una descrittiva, associando all'una i nomi di Cartesio, Leibniz e Berkeley, mentre all'altra quelli di Aristotele e Kant. Aldilà della reale plausibilità filosofica del distinguere così nettamente tra due categorie di metafisica, a mio parere Strawson intendeva identificare quel tipo di ricerca filosofica volta ad indagare la struttura cognitiva della mente umana, tesa a descriverne i processi ed i modi partendo da quanto intuitivamente possiamo conoscere di essa. Piuttosto che tentare di rivoluzionare il nostro modo di concepire noi stessi ed il nostro rapporto con la realtà esterna, si pensi alla rivoluzione attuata da Cartesio, la metafisica descrittiva presentata da Strawson è quella ricerca volta a spiegare ed a dirimere le contraddizioni insite nelle intuizioni che abbiamo su noi stessi e sul mondo.

Credo personalmente che l'intero dibattito sull'identità personale, per come si è caratterizzato da Locke in poi, possa essere definito come un dibattito di metafisica descrittiva. La storia di questo dibattito infatti, a mio parere può essere riassunta nel tentativo di rendere conto consistentemente delle intuizioni di fondo che abbiamo sull'identità personale, sia considerate per loro stesse, sia per fare in modo che si associno coerentemente con le loro vicine, pensiamo al fatto che uno degli intenti di Locke nello sviluppare la sua teoria dell'identità personale era quello di dare un fondamento al suo pensiero morale ed ai concetti di ricompensa e punizione. Un'altra questione fondamentale, con la quale necessariamente il dibattito sull'identità personale ha dovuto fare i conti, è la discussione sulla definizione generale di identità.

È evidente che una teoria che si proponga di descrivere i modi e le caratteristiche del nostro poter individuare e reindividuare una stessa persona della quale si è avuta nel tempo una percezione discontinua, debba avere il suo fondamento in una più generale teoria dell'individuazione. A questo proposito sarà interessante notare nell'arco del testo come frequentemente l'identità personale venga accostata all'identità degli artefatti.

Questo avviene perché gli artefatti hanno caratteristiche tali da poter svolgere la funzione di controparti intuitive per le persone sotto alcuni punti di vista e, soprattutto, perché permettono di riformulare in maniera più immediata alcuni casi possibili ed enigmatici in cui apparentemente si hanno intuizioni contrastanti sull'identità personale. Mi sto riferendo in particolare alle descrizioni dei casi di fissione, scambio di cervelli, eccetera che si avrà modo di incontrare più avanti ed al loro consueto accostamento con il puzzle sull'identità sulla nave di Teseo presentato per la prima volta da Hobbes o con il problema del ponte di Santa Trinità a Firenze distrutto dai nazisti e successivamente ricostruito che invece troviamo in Shoemaker (1963).

Lo sforzo di molti ricercatori dell'ultimo secolo che si sono interrogati riguardo al concetto di persona ed alla sua identificazione, è stato quello di distinguere tra caratteristiche fondamentali e secondarie di questa e, conseguentemente, di teorizzare l'esistenza di criteri necessari e/o sufficienti per stabilire se una persona al tempo t sia identica o meno ad una persona al tempo t' . Questa categoria di visioni viene appunto definita Criterialismo per l'identità personale, il quale si suddivide tra i sostenitori del criterio fisico, ovvero quelli che ritengono che l'unico reale criterio necessario e sufficiente per l'identità personale sia di natura fisica, ed i sostenitori del criterialismo psicologico, i quali invece, sulle orme di Locke, considerano la continuità/connessione mnemonica e psicologica un criterio, almeno sufficiente, per stabilire l'identità personale. A questi si è contrapposto uno schieramento di pensatori che invece difende una visione "semplice" della persona e dell'identità personale, sostenendo che questa non possa essere ridotta a criteri necessari e sufficienti, ma che invece sia un concetto fondamentale e non ulteriormente scomponibile della nostra cognizione. Questa posizione, che traspare per la prima volta dalle obiezioni che Thomas Reid e Joseph Butler rivolgono alla teoria di Locke, verrà poi difesa nella seconda metà del novecento da diversi filosofi come R. Chisholm, E. J. Lowe e soprattutto P.F. Strawson.

Il dibattito tra queste due posizioni contrastanti non è nello specifico l'oggetto di questo testo. Piuttosto in questa tesi mi propongo, dopo aver presentato sinteticamente la storia del dibattito sull'identità personale per come si è sviluppata soprattutto nella seconda metà del novecento, di mostrare e di criticare alcune conclusioni a cui sono arrivati alcuni sostenitori del criterialismo psicologico conseguenti allo svilupparsi delle loro teorie ed all'applicazione di queste ai casi enigmatici citati sopra. Queste situazioni, denominate appunto *puzzle cases* dell'identità, hanno portato infatti diversi sostenitori del criterialismo psicologico a rivoluzionare il concetto stesso di identità rifiutando l'*only x and y principle*, ovvero il principio per cui l'identità di un oggetto qualsiasi con un altro temporalmente successivo deve essere determinata esclusivamente in base a fattori intrinseci ai due oggetti stessi. Attraverso questa strada sono entrate a far parte del dibattito posizioni che, pur continuando a sostenere di voler rendere conto delle nostre intuizioni di base riguardo all'identità personale, sono arrivate a conclusioni fortemente controintuitive, prima tra tutte quella di Parfit.

A mio parere la difesa di un criterialismo, in particolar modo di quello psicologico, porta, nel momento in cui ci si confronta con il problema della fissione di un cervello umano ad esempio, a dover abbracciare posizioni fortemente controintuitive riguardo all'identità personale e soprattutto a dover utilizzare un concetto di identità privo dell'*only x and y principle*, cosa che io ritengo di per sé difficilmente difendibile, con la conseguente ammissione della plausibilità dell'esistenza di più candidati rivali per

l'identità con un soggetto gerarchizzabili in base al *grado* di identità o continuità. Queste ipotesi vengono spesso difese ed avvalorate dai loro sostenitori attraverso l'affermazione di una loro maggiore potenza esplicativa rispetto alle formulazioni tradizionali, caratteristica che permetterebbe a queste teorie di rendere più esaustivamente conto delle intuizioni contrastanti che abbiamo riguardo ai casi enigmatici sull'identità. Molti oppositori di queste posizioni però hanno efficacemente attaccato la pretesa di queste Teorie del miglior candidato, come le chiamerò in seguito, di essere in grado di spiegare e di rendere conto delle nostre intuizioni riguardo ad un ventaglio più ampio di situazioni possibili, spogliandole di un forte argomento a loro favore.

Particolarmente interessante ai fini di questo testo è la proposta teorica di Noonan, che viene da lui espressa in *Personal Identity* (2003) ed in *Personal Identity And Its Perplexities* in Gasser & Stefan (2012). Noonan tenta di formulare una visione che, pur rimanendo un criterialismo psicologico, riesca consistentemente a mantenere saldo nella sua intuitiva validità l'*only x and y principle*. Nel perseguire i suoi intenti Noonan si contrappone tenacemente ad altri esponenti della sua corrente criterialista, come Parfit e Nozick che, per difendere la loro posizione, di fronte ai problemi derivanti dai *puzzle cases* sull'identità hanno escluso dal concetto di identità l'*only x and y principle* e sostenuto quindi una Teoria del miglior candidato. Oltre alle obiezioni rivolte da Noonan a queste teorie, trovo che sia fondamentale ed ancora insuperata l'obiezione formulata da Gareth Evans nel suo articolo lungo appena una pagina *Can There Be Vague Objects?* (1978) pubblicato in *Analysis* 38.4, 1978, p. 208 nel quale afferma e prova formalmente che la relazione stessa di identità tra due oggetti non può essere vaga o indeterminata. La vaghezza di alcuni enunciati di identità non risiede negli oggetti stessi, ma nel nostro linguaggio e nel nostro modo di riferirci ad essi.

La tesi che argomenterò nell'ultimo capitolo riguardo la teoria di Noonan però sarà a sostegno del fatto che, nonostante abbia il merito dell'aver mostrato in *Personal identity* le contraddizioni delle Teorie del miglior candidato e ne abbia puntualmente criticato la pretesa di riuscire a rendere conto più esaustivamente delle nostre intuizioni riguardo a casi particolari ed enigmatici sull'identità personale, non sia riuscito nell'intento di formulare una teoria che coerentemente si caratterizzasse come un criterialismo psicologico e che contemporaneamente mantenesse valido l'*only x and y principle*. In particolare a mio parere il modo in cui Noonan supera l'obiezione formulata da Bernard Williams in *The Self And The Future* (1970) è discutibile e fallisce nel suo intento di smantellare la sua forza argomentativa diretta contro i criterialismi psicologici e, se osservata attentamente, contro i criterialismi in generale.

Mentre i primi due capitoli avranno semplicemente uno scopo introduttivo a quelli che saranno i problemi affrontati, il terzo ed il quarto capitolo rappresentano la parte concettualmente più densa del testo, in cui verranno affrontate e criticate prima le Teorie del miglior candidato e successivamente la proposta alternativa di Noonan. Lungi dall'aver la pretesa di argomentare e difendere una posizione originale riguardo ad i temi trattati, scopo di questo testo è solo sostenere l'implausibilità delle posizioni sull'identità personale che rifiutano l'*only x and y principle* e mostrare come lo svilupparsi di questo tipo di visioni sia una conseguenza, se non necessaria almeno molto difficilmente evitabile, dell'applicazione del criterialismo psicologico ai casi enigmatici sull'identità personale. Conseguenza che nemmeno l'originale proposta di

Noonan è per ora riuscita ad evitare.

1. CENNI STORICI.

In questo capitolo cercherò brevemente di inquadrare dal punto di vista storico il problema dell'identità personale prendendo principalmente in considerazione quanto scritto da John Locke nel suo *An Essay Concerning Human Understanding* (1689, Book II, Cap. 27) e le obiezioni a lui rivolte da Leibniz, Thomas Reid e Joseph Butler.

Importante è tenere in considerazione il fatto che questo dibattito nasce come costola della perenne discussione sulla definizione di Persona, di cui troviamo già formulazioni in Platone ed Aristotele. Uno dei principali intenti di Locke nello scrivere il suo *Essay* era infatti quello di fornire una alternativa al dualismo di matrice Cartesiana, teoria che aveva recentemente rivoluzionato la concezione della Persona dividendola in *Res Cogitans* e *Res Extensa*. Cartesio aveva così distinto ontologicamente mente e corpo e identificato la Persona con la sola *Res Cogitans*, immateriale e immortale. Nonostante abbia scelto, per ragioni di spazio, di non trattare esaustivamente la storia di questo dibattito è fondamentale tenere a mente questa premessa per capire quali fossero gli intenti di Locke nello stilare la sua teoria sull'Identità Personale. Oltre a voler fornire una concezione utile in materia forense infatti, uno dei principali problemi che il filosofo ha dovuto affrontare è stato quello di formulare una teoria della Persona che, similmente al dualismo Cartesiano, fosse coerente con la dottrina cristiana dell'immortalità dell'anima, della resurrezione e del giudizio finale.

1.1 Locke.

Come ho già accennato, tra i principali obiettivi di Locke nel presentare la sua teoria sull'Identità Personale troviamo quello di fornire una concezione della Persona che renda conto della possibilità della resurrezione e dell'immortalità dell'anima, senza per questo dover sottostare ad un'ontologia dualistica; quello di fornire una visione che renda conto del fatto che la conoscenza che abbiamo della nostra Identità attraverso il tempo è invulnerabile rispetto a qualsiasi obiezione scettica e inoltre, che renda conto del fatto che la nostra identità è qualcosa che per noi ha importanza, ovvero del fatto che ciò che abbiamo fatto o sofferto e ciò che faremo o soffriremo non può esserci indifferente.

Locke alla fine della sua argomentazione arriverà a concludere che l'identità di coscienza è ciò che determina l'Identità Personale, delineando una teoria che in termini moderni viene considerata un Criterialismo Psicologico. Per arrivare a questa conclusione il filosofo inglese opera una distinzione ontologica tra *Persona*, *Sostanza pensante* e *Uomo* separando di conseguenza e definendo singolarmente i loro criteri di

identità.

Una delle prime preoccupazioni dell'autore nel presentare la sua teoria è quella di fornire quello che chiama il *Principium Individuationis*. L'argomentazione di Locke muove dalla premessa che le idee di identità e diversità derivino dalla comparazione di qualcosa considerato come esistente in un determinato spazio e tempo con se stesso esistente ad un tempo differente, da ciò arriva rapidamente ad affermare che una cosa non può avere due inizi di esistenza e che due cose differenti non possono avere lo stesso. Quello che Locke sta in questo modo cercando di fornire è un criterio di identità necessario e sufficiente e infatti nel giro di poche righe arriva ad affermare il suo *Principium Individuationis*:

Existence itself, which determines a being of any sort to a particular time and place, incommunicable to two things of the same kind. (Essay, II. Cap 27.)

Che potrebbe essere parzialmente formalizzato in: x è identico a y se e solo se x e y sono dello stesso tipo e ad un certo tempo t occupano lo stesso spazio.

Nonostante Locke fornisca questa teoria generale dell'Identificazione prima di introdurre il problema dell'identità personale, come se volesse gettarne le basi, quello che arriverà ad affermare riguardo alla persona sarà incompatibile con il principio appena delineato, l'autore stesso infatti insisterà sul fatto che nel momento in cui l'identità di coscienza si interrompe, una persona potrà avere due inizi di esistenza e questo è ciò in cui consiste la possibilità della resurrezione.

A prima vista questa sembra una palese contraddizione; Locke infatti in un passaggio precedente sostiene che le uniche sostanze di cui possediamo un'idea sono: Dio, Intelligenze finite e Corpi, mentre tutte le altre cose non sono che Modi o Relazioni. Quello che Locke ha individuato con il suo *Principium Individuationis* quindi, più che un criterio generale di identità, sembra essere più plausibilmente un criterio sostanziale di identità. A questo punto però come è possibile che le persone, da lui definite come:

A thinking intelligent being that has reason and reflection and can consider itself as itself, the same thinking thing in different times and places. (Essays II, Cap 27)

non rientrino nella categoria di *sostanze*? Come già accennato sopra, la teoria di Locke si basa sulla tripartizione ontologica dell'essere umano in *Sostanza Pensante*, *Uomo* e *Persona* e sulla conseguente distinzione dei criteri di identità di questi tre tipi. Questo significa che nel momento in cui il filosofo inglese in questa discussione utilizza il termine *sostanza* non si riferisce semplicemente a cose materiali, categoria entro la quale rientrerebbero sia *Uomo* inteso come animale umano, sia *Persona*, ma si riferisce ad una categoria più ristretta che Alston e Bennett (1988) hanno identificato con:

Thing-like item that is quantified over at a basic level of one's ontology. (pp.39)

Secondo questa formulazione, quindi, attraverso il termine *sostanza* Locke si riferirebbe, nel capitolo sull'identità, ai costituenti fondamentali della realtà, ovvero a quegli oggetti dai quali dipende l'esistenza di tutti gli altri. In questo senso ad esempio

Dio e gli atomi materiali sono *sostanze*, mentre le *persone* non appartenerebbero a questa categoria in quanto l'intera loro storia può in linea di principio essere raccontata senza mai doversi riferire esplicitamente a loro ma solo alle *sostanze pensanti*, siano esse materiali o immateriali, che le costituiscono. La contraddizione sopra riportata diventa in questo modo solo apparente, in quanto Locke in questi termini può consistentemente affermare, contrariamente a quanto afferma Cartesio, che l'identità personale può non comportare l'identità della *sostanza pensante*.

Dopo aver discusso l'identità delle *sostanze* Locke procede nella sua argomentazione prendendo in considerazione l'identità degli organismi. L'identità di uno di questi non può dipendere dall'identità delle particelle di materia che lo costituiscono in quanto nel corso della sua vita uno stesso organismo rinnova più volte ogni particella materiale che lo costituisce. Prendendo ad esempio l'identità di una pianta Locke scrive:

That being than one plant which has such an organization of parts in one coherent body, partaking of one common life, it continues to be the same plant as long as it partakes the same life.

(Essay II. Cap 27)

Lo stesso vale nel caso nell'identità degli animali o degli uomini, Locke infatti insiste sul fatto che ciò in cui consiste l'identità attraverso il tempo di un organismo vivente è la partecipazione alla stessa vita che si perpetua rinnovando particelle di materia continuamente organizzate nello stesso corpo. Dopo aver definito in questo modo l'identità degli organismi viventi, definisce analogamente l'identità personale sostenendo che così come l'identità di un organismo consiste nell'identità della vita, l'identità della persona consiste nell'identità della coscienza; e similmente così come l'identità della vita può attuarsi attraverso diverse organizzazioni di sostanze materiali, allo stesso modo l'identità della persona può persistere nonostante si succedano diverse sostanze pensanti.

Harold Noonan (2003. Cap 2) individua due linee di pensiero che, coerentemente con gli intenti di Locke, hanno plausibilmente portato l'autore a questa conclusione. Uno di questi intenti era quello di fornire una concezione dell'identità personale che rendesse conto della possibilità della resurrezione e dell'immortalità, senza sottostare ad un'ontologia dualistica di tipo Cartesiano. Se Locke avesse adottato una strategia materialista difficilmente sarebbe riuscito ad assolvere il compito che si era proposto; concepita in quest'ottica, infatti, l'identità personale non può consistere semplicemente nell'identità delle sostanze materiali che costituiscono un corpo, in quanto queste mutano e si succedono dinamicamente nell'arco di una stessa vita. L'identità personale risulterebbe essere quindi soltanto un caso speciale dell'identità animale, preservata dall'identità della vita, la quale però plausibilmente richiede che sia mantenuta la continuità spazio-temporale, che è esattamente il requisito di cui Locke sta cercando di fare a meno per poter ammettere la possibilità della resurrezione. Secondo questa concezione, inoltre, verrebbe a mancare la differenza tra identità dell'animale umano e identità della persona, distinzione che Locke spesso sottolinea come fondamentale per poter rendere conto seriamente delle implicazioni forensi del concetto di persona, che verrebbero invece salvaguardate da una concezione dell'identità personale fondata su una continuità psicologica come quella che Locke ha formulato.

La seconda linea di pensiero individuata da Noonan, di tipo epistemologico, muove dall'esplicito tentativo del filosofo inglese di rendere conto della nostra certezza dell'essere stati i fautori di determinate azioni e pensieri passati in quanto le stesse persone che stanno agendo nel presente e del fatto che questa conoscenza non può esserci indifferente. Locke sostiene che, nel ricordare una qualche precedente azione, mentre lecitamente siamo certi di essere quella stessa persona che ha compiuto l'azione, non possiamo essere certi di essere la stessa sostanza pensante, sia essa materiale o immateriale. Questa distinzione viene posta dall'autore sulla base del fatto che il nostro flusso di coscienza viene costantemente interrotto, sia drasticamente dal sonno o da qualche trauma, sia semplicemente per il fatto che non siamo mai consciamente in possesso dell'intera registrazione mnemonica di ogni nostra passata azione ma, al contrario, questa è episodica e spesso ne dimentichiamo delle parti di modo che "*we lose sight of our past selves*" (*Essays II, Cap 27, 10*). Locke sostiene quindi che, non potendo noi essere certi di essere la stessa sostanza pensante che siamo stati in passato, non possiamo fondare su queste ultime il nostro criterio di identità personale. Secondo questa teoria, quindi, l'unica cosa di cui possiamo essere certi nel ricordare una nostra precedente azione è l'identità di coscienza, per Locke infatti non possiamo essere certi neanche dell'identità della vita. Il filosofo inglese pone il problema nei termini di una "certezza" che siamo o meno autorizzati ad avere riguardo a quali siano le nostre fondamenta epistemologiche nel considerarci la stessa persona attraverso il tempo, da ciò ne deriva che le uniche risposte ammesse in quest'ottica siano quelle che fanno riferimento a quanto è aperto alla mia introspezione. Conclusione anche di questa linea di pensiero individuata da Noonan è che l'unica cosa di cui possiamo essere certi, e quindi l'unico candidato a criterio di identità personale, è l'identità di coscienza.

Con queste argomentazioni Locke sta cercando di rendere conto del fatto che la nostra identità di persone attraverso il tempo non è semplicemente qualcosa di teoreticamente importante per noi, ma lo è ugualmente a livello pratico. Locke sostiene che nonostante per una qualche improbabile coincidenza noi potremmo essere composti in parte dalle stesse sostanze materiali che composero un tempo Nestore, questo non implica che dovremmo avere un qualsiasi interesse per le passate azioni di questo personaggio della mitologia greca e, nel caso in cui venissimo a conoscenza di condividere con Nestore alcune sostanze materiali, non avrebbe comunque alcun senso attribuirsi e valutare come proprie le sue passate azione né similamente avere per questo un interesse specifico nella sua vita. Analogamente secondo Locke noi potremmo aver condiviso con Nestore una sostanza pensante per un certo periodo di tempo, ma questo è consistente con la possibilità che noi non condividiamo alcun tratto psicologico con lui né abbiamo coscienza delle sue passate azioni come se fossero le nostre.

La differenza cruciale che Locke voleva tracciare con queste argomentazioni tra l'identità della sostanza, sia essa materiale o pensante, e l'identità di coscienza è il fatto che mentre la mia conoscenza di condividere con qualche persona passata o futura la stessa sostanza materiale o pensante può lasciarmi indifferente riguardo alle azioni di questa, nel momento in cui io diventi conscio delle azioni di questa, non posso rimanere indifferente ma anzi sono costretto a considerare quelle azioni come mie e ad imputarmele in senso sia pratico che morale. Sono costretto in conclusione a considerare la persona che ha compiuto quelle azioni come la stessa persona che sono in

questo momento.

Con queste ultime considerazioni siamo arrivati ad un punto cruciale della teoria lockeana, ovvero la tesi che *persona*, in ultima analisi, sia un termine forense, il quale si usa per definire esseri pensanti capaci di formulare e sottostare a delle leggi e che viene utilizzato per attribuire i meriti e le colpe conseguenti alle azioni di questi. Locke sostiene che il concetto di identità personale sia il fondamento dei concetti di *ricompensa* e *punizione*, in quanto è impensabile che io sia indifferente alle conseguenze delle azioni che concepisco come mie. Similmente ognuno è preoccupato della felicità o miseria considerate come proprie in relazione ad una continuità di coscienza.

Il requisito della continuità di coscienza nella identificazione di una stessa persona, passibile quindi di essere punita o ricompensata in rapporto alle azioni da lei compiute, ha delle conseguenze che sembrano essere difficilmente conciliabili con il piano giuridico. Sembrerebbe infatti che una persona possa in quest'ottica essere punita solo per aver commesso azioni che lei possa ricordare, in questi termini quindi esisterebbe la possibilità che un criminale riesca ad evitare la sua punizione semplicemente inscenando una totale amnesia. Locke si confronta con questo problema ipotizzando il caso di una persona che commette un qualche crimine mentre è ubriaca e, una volta tornata sobria, non si ricordi di averlo commesso. Secondo la visione di Locke punire il sobrio per un'azione commessa in stato di ubriachezza che non è in grado di ricordare sarebbe ingiusto quanto punire una persona per qualcosa da lei commessa durante un momentaneo attacco di pazzia, sarebbe come punire una persona per quanto commesso da un'altra. Ad ogni modo Locke era conscio del fatto che, similmente alle corti moderne, le corti della sua epoca erano solite punire le persone per quanto commesso da ubriache nonostante queste non riuscissero a ricordare le loro azioni. In un passaggio molto criticato del suo *Essay*, l'autore cerca di riconciliare la sua teoria con i procedimenti giuridici dicendo che l'apparente contraddizione tra l'una e la pratica degli altri era causata dal fatto che nella situazione possibile consistente in un crimine commesso in stato di amnesia da ubriachezza, non possono esserci prove reali dell'effettiva continuità o meno della coscienza della persona. Questo tentativo di riconciliazione, ad ogni modo, è fallimentare in quanto la motivazione giuridica sottostante alla differente considerazione dei casi di amnesia da pazzia o da ubriachezza risiede nel fatto che l'ubriachezza è uno stato volontariamente indotto. Il problema fondamentale della trasposizione forense della teoria di Locke sta nel suo insistere nel considerare la continuità di coscienza come condizione non solo *necessaria*, ma anche *sufficiente* per la responsabilizzazione di una persona riguardo alle sue azioni, mentre giuridicamente anche nell'Inghilterra di Locke erano previste situazioni valutate come scuse od attenuanti anche in casi in cui non potevano esserci dubbi riguardo all'identità personale.

Queste ultime considerazioni non colpiscono direttamente la teoria di Locke ma solo il suo tentativo di utilizzarla in ambito forense, ad ogni modo il tentativo di Locke di proporre la sua teoria dell'identità personale come fondamento dei concetti morali di *punizione* e *ricompensa* ed il fatto che l'identità di coscienza venga definita condizione *necessaria* e *sufficiente* per l'identità personale, verranno utilizzati come punti deboli sui quali far leva per formulare obiezioni a questa teoria.

1.2 Leibniz.

Tra le prime trattazioni di Leibniz del problema dell'identità personale particolarmente rilevante è quella contenuta nella sezione xxxiv del suo *Discourse De Métaphysique* (1686) intitolata "Sulla differenza degli spiriti dalle altre sostanze, anime o forme sostanziali, e sul fatto che l'immortalità che è richiesta include la memoria". In questa sezione la teoria delineata da Leibniz non si discosta molto da quella lockeana che non aveva ancora avuto modo di studiare, anche Leibniz infatti concepisce "persona" come un termine forense ed è perciò preoccupato di fornire una concezione dell'identità personale che renda conto del perché essere la stessa persona in due tempi differenti abbia importanza e sia qualcosa che non può esserci indifferente. È quindi portato, per gli stessi motivi di Locke, ad assegnare un ruolo fondamentale alla continuità della memoria ed al mantenimento delle proprie conoscenze, condizioni necessarie per rimanere la stessa persona dal punto di vista morale.

Leibniz insiste più volte sul fatto che l'anima umana immortale non solo persiste come se stessa metafisicamente come sostanza, ma anche moralmente come continuità di memoria e conoscenza, in quanto è sulla base di questa che è passibile di *punizione* o *ricompensa*. Leibniz sostiene che le sostanze materiali o immateriali, come l'anima degli animali o i costituenti primi della materia, non possano perire. Ciò che distingue queste sostanze dagli spiriti intelligenti, come l'anima dell'uomo, è la loro incapacità di riflettere attivamente su loro stesse e di utilizzare concetti logici attraverso i quali noi siamo in grado di scoprire verità universali. Secondo Leibniz infatti è grazie a questa nostra capacità di riflettere su quello che noi stessi siamo e di dire "Io" che la nostra anima persiste non solo metafisicamente, ma anche moralmente ed è perciò immortale e non soltanto imperitura.

Il concetto di identità personale anche in questa trattazione viene utilizzato come fondamento della responsabilità morale e "persona" come un termine forense che attribuisce meriti e colpe alle azioni di qualcuno. In un esperimento mentale molto simile a quello di Locke in cui noi casualmente veniamo a condividere lo stesso spirito immateriale che fu un tempo quello di Nestore, Leibniz immagina una situazione in cui ad una persona viene detto che il giorno successivo al suo risveglio sarà improvvisamente re della Cina, ma non conserverà alcuna memoria né conoscenza della sua vita passata, sostenendo che date queste condizioni questa persona non dovrebbe plausibilmente avere alcun desiderio che ciò accada. Questo esperimento mentale è molto simile a quello Lockeano dal quale si differenzia solo per il fatto che prende in considerazione l'attitudine mentale di una persona verso il proprio futuro, mentre Locke dirigeva l'attenzione del suo soggetto sperimentale verso il proprio passato; in entrambi i casi la conclusione alla quale arrivano i due filosofi è la stessa, ovvero che la continuità di memoria è condizione necessaria per l'identità personale.

Mentre però Locke sostiene che questa condizione sia anche sufficiente a definire il concetto di identità personale, Leibniz fin dalla sua prima esposizione del problema ritiene che questa debba essere affiancata all'identità sostanziale della persona. Per evitare la possibilità, prevista nella teoria di Locke, che ad una sostanza pensante che ad un certo tempo si qualifica come una determinata persona, vengano annullate tutte le sue passate memorie e conoscenze di modo tale che ad un tempo differente questa

sostanza pensante risulti per il criterio della continuità psicologica una persona diversa, Leibniz fa affidamento al volere di Dio, del quale dice che senza dubbio ha ordinato tutte le cose in modo che gli Spiriti Immateriali conservino non solo la loro sostanza, indispensabile alla loro persistenza, ma anche le loro qualità morali indispensabili perché questa persistenza diventi immortalità. In questo modo Leibniz ha escluso anche la possibilità della realizzazione effettiva del suo esperimento mentale del re della Cina, anche se solo su un piano teologico.

Dopo aver letto il testo lockeano però Leibniz, nei suoi *Nouveaux Essais Sur L'Entendement Humain* (1705), rivolge alcune obiezioni al filosofo inglese che si distaccano sensibilmente dal pensiero espresso in precedenza. Non solo infatti Leibniz non abbandona il requisito della identità sostanziale, ma in un passaggio degli *Essais* sembra distaccarsi ulteriormente da Locke più di quanto lo fosse già nel *Discourse*, arrivando a sostenere che la memoria non rappresenta più nemmeno un criterio necessario per l'identità personale. Leibniz obietta alla teoria di Locke che è innegabile che una persona sia la stessa persona che neonata è stata cullata dalla madre nonostante di ciò non abbia alcun ricordo, e perché si mantenga l'identità morale è sufficiente che esista una connessione di coscienza tra stati mentali tra loro prossimi. Anche nel caso di un'interruzione di coscienza dovuta ad esempio ad un'amnesia le testimonianze ed i racconti di altre persone possono riempire le lacune formatesi nella memoria della persona, di modo tale che questa possa, nel caso, essere punita per qualcosa da lei commesso ma dimenticato per causa dell'amnesia. In un passaggio successivo Leibniz si spinge ancora più lontano da quanto sostenuto nel *Discourse*, arrivando a sostenere che anche nel caso in cui una persona subisca una totale perdita di memoria, questa può comunque apprendere dagli altri quanto concerne il modo in cui aveva condotto la propria vita prima dell'amnesia, mantenendo così la propria identità morale.

Leibniz comunque non arriverà a negare che l'identità personale implichi una qualche forma di connessione psicologica, sosterrà piuttosto che uno spirito immateriale non possa essere separato completamente da tutte le percezioni della sua passata esistenza perché tenderà sempre a mantenere impressioni di quanto gli è accaduto, ed è in base a queste che è in grado di compiere previsioni su quanto gli accadrà. Secondo Leibniz è questa continua interconnessione di percezioni che rende una persona la stessa in due tempi diversi, mentre è la coscienza di questa continuità con gli stati mentali passati a fondare l'identità morale.

Leibniz infine rivolge un'obiezione alla teoria di Locke per formulare la quale si serve di un altro esperimento mentale estremamente simile a quello che, circa due secoli e mezzo dopo, verrà chiamato da Bernard Williams il *Reduplication Argument*. Leibniz nei *New Essays on Human Understanding* (pp. 245) ipotizza l'esistenza di un mondo perfettamente identico al nostro in cui si trovano i doppiati fisici e mentali perfetti di tutte le persone presenti nel nostro, ipotizza inoltre che Dio possa a piacimento scambiare le menti o le intere controparti senza che queste se ne possano accorgere. In un caso come questo la teoria di Locke, secondo la quale è l'identità di coscienza a determinare l'identità personale, risulta essere incapace di distinguere tra le controparti dei due mondi e porta, anzi, a considerare le controparti come la stessa persona, nonostante queste possano essere considerate come due diverse da Dio.

La diversità nel pensiero di Leibniz, per come viene formulato negli *Essais*, rispetto a quello di Locke risiede nel fatto che il requisito della connessione psicologica non

viene considerato incompatibile con quello dell'identità sostanziale ma, anzi, ne è fondamento ed è proprio in virtù della continuità delle percezioni che uno spirito rimane la stessa sostanza nel tempo. Criterio necessario e sufficiente quindi in quest'ottica è l'identità sostanziale dello spirito immateriale, inteso da Leibniz come una monade attiva la cui identità attraverso il tempo è fondata nella connessione psicologica tra i suoi stati mentali. La continuità psicologica quindi non è più in questa formulazione un criterio per l'identità personale ma rimane il fondamento del nostro essere agenti morali e, come tali, individuati attraverso uno spirito immateriale che non solo persiste ma è immortale.

1.3 Joseph Butler e Thomas Reid.

Tra le innumerevoli obiezioni che Butler e Reid hanno rivolto alla teoria di Locke solo alcune si sono dimostrate significative, in questo testo in particolare ne prenderò in considerazione tre. La differenza fondamentale tra le obiezioni che sono state rivolte a Locke da Leibniz e quelle invece provenienti da Butler e Reid risiede nel fatto che mentre la prospettiva di partenza di Leibniz aveva diversi punti di contatto con quella lockeana, i due studiosi britannici attraverso le loro obiezioni rivelano una visione di fondo profondamente contrastante con quella del loro oppositore. Mentre Locke viene considerato il primo propositore di una di quelle visioni che in termini moderni vengono chiamate *Criterialismi psicologici* sull'identità personale, quello che traspare dalle le obiezioni di Butler e Reid viene considerato come il primo tentativo di argomentare a favore di quella che, sempre in termini moderni, verrebbe chiamata una *Simple View*. La differenza fondamentale tra qualsiasi tipo di criterialismo riguardo all'identità personale e quanto invece sostenuto dai difensori delle *simple views*, è il fatto che, mentre il criterialismo si propone di ridurre l'identità personale a qualcosa di più fondamentale che ne è condizione necessaria e sufficiente, come ad esempio la continuità psicologica, secondo i sostenitori delle *simple views* i concetti di persona e di identità personale sono già di per se stessi elementi fondamentali della nostra cognizione e non sono perciò ulteriormente riducibili.

La prima e più facilmente eludibile obiezione tra quelle qui prese in considerazione riguarda il fatto che, per come viene presentata, la teoria di Locke non rispetta la transitività del concetto di identità, ovvero il fatto che se a al tempo t' è identico a b al tempo t'' e b al tempo t'' è identico a c al tempo t''' , allora a al tempo t' è identico a c al tempo t''' . Il vescovo Butler esprime le sue perplessità riguardo alla teoria di Locke mettendo in luce l'assurdità del fatto che, secondo quanto da lui espresso una persona non è esistita un singolo momento né ha compiuto una sola azione se non quelle che è in grado di ricordare. Reid invece più esplicitamente propone il diventato celebre paradosso del "Gallant Officer", in particolare scrive:

"Suppose a brave officer to have been flogged when a boy at school for robbing an orchard, to have taken a standard from an enemy in his first campaign, and to have been made a general in his advanced life; suppose also... that, when he took the

standard, he was conscious of his having been flogged at school, and that, when made a general, he was conscious of his taking the standard, but he had absolutely lost the consciousness of his flogging...it follows, from Mr. Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follow, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach as far back as his flogging; therefore according to Mr. Locke's doctrine he is not the same person who was flogged. Therefore the general is, and at the same time is not, the same person with him who was flogged at school. (Reid 1941:213)

Nonostante questa obiezione abbia pienamente ragion d'essere per come è stata formulata la teoria di Locke nel suo *Essay*, nondimeno risulta essere piuttosto facilmente superabile; è sufficiente infatti sostituire, come aveva già intuito Leibniz nei *Nouveaux Essais*, il requisito della continuità, con quello della connessione mnemonica, di modo tale che per mantenere l'identità personale sia sufficiente che siano tra loro connessi solo gli stati mentali tra loro prossimi così concatenati. Questa soluzione avrebbe cavato d'impaccio Locke anche per quanto riguarda le obiezioni a lui rivolte in merito al suo pensiero sull'affidabilità della responsabilità morale nei casi di amnesia da ubriachezza che abbiamo già precedentemente citato.

Butler e Reid accusano anche la teoria lockeana di aver in realtà fornito una definizione circolare dell'identità personale. Dal loro punto di vista la continuità del flusso di coscienza viene necessariamente presupposta dal concetto di identità personale e non può quindi costituirne un criterio. Questa obiezione si rivolge in particolare a tutti i tentativi di definire l'identità personale in termini mnemonici, in quanto nell'analisi della nozione di memoria, se intendiamo distinguere tra memoria veridica e non, necessariamente dobbiamo far ricorso al concetto di identità personale. Questa obiezione viene rivolta ai sostenitori dei criterialismi psicologici ancora oggi ed ha portato alcuni di essi a tentare di definire la nozione di memoria facendo a meno di utilizzare il concetto di identità personale, a cercare di definire una nozione diversa di *quasi-memoria*, termine coniato da Sidney Shoemaker (1984) che differisse dalla nozione di memoria solo per il suo non presupporre il concetto di identità, oppure a superare questa difficoltà analizzando il concetto di identità personale in termini funzionalistici, come ad esempio hanno cercato di fare Shoemaker (2004) o Agar (2003).

Ad ogni modo Butler e Reid quando rivolgevano la loro obiezione alla teoria di Locke probabilmente non immaginavano che questa sarebbe stata poi ripresa da altri e ulteriormente raffinata. Quello che sembrano intendere i due filosofi pare essere semplicemente che non è possibile definire che cosa sia l'identità personale, facendo riferimento al modo in cui siamo a conoscenza che si dia l'identità personale e, se la cosa è messa in questi termini, Locke sembra poter rispondere all'obiezione. Nonostante la nozione di coscienza che utilizza nella sua teoria sia una forma di conoscenza il contenuto di questa non è l'identità personale ed i loro possessori non sono *persone* ma *sostanze pensanti*; avendo tenute distinte le nozioni di persona e sostanza pensante Locke sembra essere immune all'obiezione di circolarità in quanto la relazione nei

termini dei quali definisce l'identità personale, ovvero identità di coscienza, non è una relazione tra persone ma tra sostanze pensanti. A prima vista questa soluzione sembra resistere all'obiezione di circolarità. Se però consideriamo più attentamente la sua teoria notiamo che Locke non distingue le proprietà psicologiche possedute dalle persone da quelle possedute dalle sostanze pensanti, di modo che sembra attribuire alle une quelle delle altre e viceversa, attribuendo quindi ad entrambe anche la coscienza. Sembra difficile quindi che la teoria di Locke si dimostri in grado di distinguere consistentemente le nozioni di persona e di sostanza pensante, e poi di farlo in modo tale che venga mantenuta coerentemente la separazione teoretica tra l'identità della sostanza pensante e l'identità della persona.

La successiva obiezione di Butler e Reid è appunto la tesi secondo la quale Locke non è stato in grado di distinguere teoricamente le nozioni di Persona e di Sostanza pensante. Questa obiezione è stata riproposta più recentemente anche da Sidney Shoemaker nei termini seguenti:

“Locke defines ‘person’ as meaning ‘a thinking, intelligent being...’ If persons are thinking things and thinking things are substances then persons are substances. And if it follows from the definition of a ‘person’ that a person is a substance, it is surely self-contradictory to say that the identity of a person does not involve the identity of a substance.
(Shoemaker 1963:45-46)

A prima vista sembra che Locke possa rispondere facilmente a questa obiezione. Se infatti si tiene a mente l'interpretazione di Alston e Bennett (1988) dell'uso che Locke fa della parola *sostanza* nel capitolo ventisettesimo del suo *Essay*, ovvero quello che la definisce come un costituente fondamentale della realtà, è possibile rispondere semplicemente che, contrariamente a quanto presupposto da questa obiezione, *Being* e *Substance* non rappresentano la stessa idea. Questa risposta però risolve solo parzialmente le perplessità degli obiettori: così come se le persone sono sostanze pensanti, la stessa persona deve essere la stessa sostanza pensante, allo stesso modo se le sostanze pensanti sono persone, la stessa sostanza pensante deve essere la stessa persona. Considerando l'interpretazione della parola “sostanza” proposta da Alston e Bennett, siamo in grado di bloccare solo la prima inferenza, ovvero quella che da “x è una persona” porta a “x è una sostanza pensante” in quanto possedere le caratteristiche psicologiche che Locke attribuisce alle persone non rende necessariamente queste ultime dei costituenti fondamentali della realtà. L'inferenza che porta da “x è una sostanza pensante” a “x è una persona” non viene similmente bloccata in quanto, secondo la teoria di Locke, sono le sostanze pensanti che “pensano nelle” persone a possedere tutte le caratteristiche psicologiche che rendono le persone ciò che sono. Se però quindi tutte le caratteristiche psicologiche delle sostanze pensanti sono condivise dalle persone in cui pensano, difficilmente si potrà dimostrare che una persona non possa necessariamente essere un costituente fondamentale della realtà.

Joseph Butler, nel contrapporsi alle tesi di Locke, insiste molto sulla distinzione tra un senso popolare e largo di identità ed uno stretto e filosofico, definendo il primo come atto a servire agli scopi della vita di tutti i giorni per attenerci ai quali possiamo ad esempio parlare di uno “stesso albero”, nonostante questo non sia lo stesso in un senso stretto e filosofico. Questo secondo significato di “identità” viene definito da Butler

come applicabile solo nel caso in cui non ci sia stato alcun cambiamento in nessuna delle parti che compongono l'oggetto a cui ci stiamo riferendo. Butler utilizza questa distinzione per sostenere che, contrariamente a quanto argomentato da Locke, una stessa persona non possa sopravvivere ad un cambio di sostanza pensante, attribuendo però piuttosto dogmaticamente la sua definizione di significato stretto e filosofico di identità all'uso lockeano.

In tempi più recenti Roderick Chisholm (1976) ha ripreso la distinzione di Butler tra il significato popolare di identità e quello filosofico per riproporre ed avvalorare la tesi secondo cui Locke non è riuscito a mantenere teoreticamente separate le nozioni di *persona* e di *sostanza pensante*. Chisholm sostiene che gli oggetti mereologicamente variabili, ovvero in grado di sopravvivere e qualificarsi come identici in senso popolare nonostante una variazione nelle parti che li compongono, vanno considerati come "ontologicamente parassitari di", o "costruzioni logiche da," oggetti mereologicamente costanti che rimangono identici in senso stretto e filosofico. Secondo Chisholm nel descrivere l'universo i riferimenti agli oggetti mereologicamente variabili, come ad esempio tavoli, sedie e quant'altro, possono essere sostituiti con riferimenti ai loro costituenti fondamentali e mereologicamente costanti. Chisholm sostiene inoltre che, tra le proprietà che un oggetto mereologicamente variabile possiede in un dato momento, un loro sottoinsieme è formato da quelle proprietà che l'oggetto ha solo grazie al fatto che queste appartengono agli oggetti mereologicamente costanti che lo costituiscono. Egli definisce questo insieme come quello delle proprietà "radicate unicamente nei tempi nei quali vengono possedute"; un tavolo ad esempio pesa cinquanta chili in un dato tempo in virtù del fatto che il blocco di materia che lo compone in quel tempo pesa cinquanta chili. Applicando questa ipotesi al caso specifico delle persone, considerate come oggetti mereologicamente variabili, avremo il risultato che proprietà psicologiche come credenze o speranze come "sperare nella pioggia", che Chisholm considera radicate unicamente nel tempo in cui sono possedute, devono necessariamente appartenere alle persone solo in virtù del loro essere state prese in prestito da quelle che Locke chiama sostanze pensanti e che rappresentano in un dato momento i costituenti fondamentali delle persone. L'argomentazione di Chisholm è criticabile da diversi punti di vista, gli si può obiettare ad esempio che le proprietà psicologiche di cui parla non siano "unicamente radicate nei tempi in cui vengono possedute", oppure che queste proprietà possano essere possedute da un oggetto mereologicamente variabile solo in virtù del fatto che sono possedute dai suoi costituenti fondamentali. Ad ogni modo la descrizione proposta da Chisholm si adatta perfettamente alla teoria lockeana secondo cui le proprietà psicologiche di una persona sono da lei possedute in virtù della sostanza pensante che pensa in lei in quel dato momento. Il punto a cui intende arrivare Chisholm è semplicemente che questo non ha senso, che semplicemente non c'è ragione per ipotizzare che le proprietà psicologiche di una persona siano da lei possedute solo in virtù dell'essere possedute da qualcosa di più fondamentale da cui la persona le "prende in prestito", non c'è ragione per pensare che io spero nella pioggia solo perché qualcos'altro che è in me "spera nella pioggia". Secondo Chisholm dunque, se esistono realmente queste due cose che in questo momento "sperano nella pioggia", la prima delle quali possiede questa proprietà mentre la seconda la "prende in prestito", allora più plausibilmente io dovrei essere identificato con la prima di queste due cose e non con la seconda.

Questa obiezione non costringe Locke a rigettare completamente la sua teoria, si potrebbe infatti continuare a sostenere che il fatto che una persona pensi solo in virtù della sostanza pensante che la costituisce è semplicemente un'ipotesi teoretica valida fino al momento in cui questa rappresenta la migliore teoria filosofica della mente di cui disponiamo. Arrivati a questo punto però la sua definizione di persona: "*a thinking intelligent being that has reason and reflection...*" sembra essere applicabile anche alle sostanze pensanti, cosa che rende difficilmente sostenibile la continua insistenza di Locke sul fatto che persone e sostanze pensanti abbiano criteri di identità differenti. Per difendere la sua teoria Locke viene costretto da questa obiezione ad insistere sulla distinzione tra l'idea di Io e l'idea dell'oggetto che possiede le caratteristiche psicologiche che troviamo nella sua definizione ufficiale di persona, ovvero deve insistere sul fatto che quando una sostanza pensante pensa "Io" il referente non è lei stessa ma la persona che costituisce in quel momento. Chisholm si chiede se nei termini di questa soluzione "*If I want my dinner, does it follow that two of us want my dinner? Or does the thinking substance want its dinner and not mine?*" (1976:108). La risposta che Locke sembra essere costretto a dare a questa domanda è che la sostanza pensante che sta pensando in me vuole che io abbia la mia cena, ma l'unico modo in cui essa è in grado di avere quel pensiero è "Io voglio la mia cena".

Anche quest'ultima obiezione alla teoria di Locke non la rende inconsistente ma costringe un suo ipotetico sostenitore a fare parecchi salti mortali per riuscire a mantenerla plausibile ed è sicuramente una delle obiezioni più interessanti che un moderno sostenitore di una teoria di spirito lockeano deve necessariamente affrontare.

1.4 Conclusioni.

Questa succinta introduzione storica, lungi dal voler essere una presentazione esaustiva del dibattito "classico" sull'identità personale, aveva come unico intento quello di introdurre alcuni argomenti e tipologie di approccio al problema in funzione di quanto verrà successivamente trattato. Ho ritenuto più rilevante presentare le radici ed i problemi fondamentali di quelle teorie che oggi vengono definite "criterialismi psicologici", in quanto esse ed alcune loro conseguenze teoretiche relative al concetto stesso di identità verranno prese in considerazione più da vicino nei capitoli che seguono. Oggetto della prima parte del prossimo capitolo sarà una presentazione in termini abbastanza generali del moderno dibattito sia tra criterialisti psicologici e criterialisti fisici, sia tra criterialisti in generale e non criterialisti. Nella seconda parte invece, sempre per motivi concernenti la rilevanza agli scopi di questo testo, mi concentrerò principalmente sui criterialismi psicologici e cercherò di entrare più specificamente nel merito di alcune argomentazioni sviluppate da Sidney Shoemaker e Derek Parfit mediante l'immaginazione di particolari pratiche di chirurgia sperimentale o di possibilità tecnologiche alla *Star Trek* sulle quali si è sviluppata un' imponente letteratura.

2.

LA RICERCA DI CRITERI PER L'IDENTITÀ PERSONALE.

Nella seconda metà del novecento e in particolare dopo l'uscita di *Individuals: An Essay In Descriptive Metaphysics* (1959) di P.F. Strawson, il dibattito sul tema dell'identità personale vede contrapporsi sostenitori delle *Simple Views* o non criterialisti e sostenitori delle *Complex Views* o criterialisti, a loro volta divisi tra fisicalisti e psicologisti. Queste distinzioni iniziali ad ogni modo rappresentano solamente l'intelaiatura più superficiale del dibattito e in questo contesto intendono solamente definire preliminarmente i termini del discorso.

Senza entrare per ora nel merito delle loro argomentazioni, i difensori delle *Simple Views* tra i quali appunto P.F. Strawson hanno ripreso e riformulato, applicandole a teorie sull'identità personale più recenti, alcune argomentazioni e obiezioni che Butler e Reid avevano rivolto a Locke, come ad esempio l'obiezione di circolarità applicabile ai criteri psicologici in generale. Secondo le visioni non criterialiste della persona, delle quali nelle pagine seguenti verrà principalmente trattata la teoria formulata da Strawson, il concetto di persona assume il ruolo di un componente fondamentale della nostra cognizione, perciò non ulteriormente scomponibile: l'identità personale, di conseguenza, non può in quest'ottica essere definita attraverso criteri che la ridurrebbero, assieme al concetto di persona, a qualcosa di più basilare. Al contrario, i criterialisti difendono la tesi che la persona e l'identità personale siano definibili attraverso criteri siano essi di natura fisica o biologica, come sostenuto da Eric T. Olson, o di natura mnemonica o psicologica, come invece sostenuto in tempi più recenti in difesa delle formulazioni lockeane da S. Shoemaker e D. Parfit e quindi riducibili a elementi più fondamentali.

Tentativo di questo capitolo è quello di illustrare i nodi principali e le tesi che caratterizzano ancora oggi il dibattito sull'identità personale, in modo tale da fornire un quadro d'insieme abbastanza generale ma sufficiente per gettare le basi e permettere comprensibilmente di approfondire, nei capitoli che seguiranno, quanto è più specificamente rilevante ai fini del program. Di analisi presentato nell'*Introduzione*. Particolarmente rilevanti si riveleranno infatti alcune tesi e argomentazioni formulate da Shoemaker e Parfit nell'ambito della difesa di un criterialismo psicologico, che verranno criticate in relazione alle loro conseguenze sul concetto di identità personale.

2.1 Strawson e la persona intesa come “Basic particular”.

Il principale intento di P.F. Strawson in *Individuals* (1959) era quello di indagare i modi in cui il linguaggio naturale è in grado di riferirsi a individui, dove per individui si intende qualsiasi cosa possa avere la funzione di soggetto logico e si possa identificare riferendocisi. In questo senso quindi vengono considerati individui sia gli oggetti materiali, i suoni, le persone, eccetera, sia i numeri, le qualità e le specie. Il libro di Strawson è diviso in due parti. Nella prima intitolata “Particulars”, l’autore indaga su quale sia il ruolo dei corpi materiali e delle persone tra gli altri particolari o individui allo scopo di verificare l’esistenza di un tipo di particolari sul cui modo di riferimento e identificazione si fonda l’identificazione degli altri. Nella seconda parte invece considera il rapporto tra gli oggetti particolari e la funzione di soggetto logico. La parte per noi più interessante è naturalmente la prima e, in particolare, il terzo capitolo intitolato appunto “Persons” dove troviamo formulata la sua “Simple view” sul concetto di persona.

Il contesto di indagine di Strawson è quindi di natura linguistica e comunicativa, ciò nonostante, dato che un parlante può riferirsi ad un oggetto senza che l’ascoltatore capisca a cosa si stia riferendo e che noi possiamo anche solo pensare ad un particolare oggetto, il rapporto tra parlante ed ascoltatore non sarà direttamente rilevante per i fini di questa discussione che si propone piuttosto di capire quali siano le condizioni generali che rendono possibile l’identificazione e la reidentificazione di un individuo.

Per poter identificare un dato particolare un parlante deve essere in grado di discriminarlo dagli altri. Nel caso più semplice e secondo Strawson fondamentale, l’oggetto in questione è presente ai sensi del parlante e questi può comodamente identificarlo attraverso l’uso di dimostrativi come “quello” o “questo”. Nel caso in cui il particolare non sia presente ai sensi del parlante invece, l’identificazione di questo necessariamente deve avvenire attraverso l’uso di descrizioni. Una stessa descrizione può però applicarsi allo stesso modo a più oggetti nell’universo facendo così fallire in linea di principio il tentativo di singolarizzazione del particolare, questo problema per Strawson è solamente apparente e si mostra come tale nel momento in cui consideriamo il fatto che la nostra conoscenza dei particolari si inserisce all’interno di un sistema spazio-temporalmente definito all’interno del quale siamo coscienti del nostro punto di vista di riferimento. Anche descrizioni come “il primo uomo nato in acqua”, le quali sembrano rivendicare il loro carattere puramente identificativo e la loro indipendenza da dimostrativi o da un condiviso sistema spazio-temporale di particolari, sollevano secondo Strawson un problema solo apparente, il metodo fondamentale attraverso il quale abbiamo conoscenza di un particolare all’interno della nostra cornice spazio-temporale condivisa è la sua identificazione mediante l’uso di dimostrativi, qualsiasi oggetto venga invece collocato da una descrizione puramente identificativa come quella sopra citata al di fuori del nostro sistema di riferimento, viene per Strawson semplicemente tagliato fuori dal poter avere un qualsiasi ruolo nelle nostre conoscenze.

Altra condizione fondamentale perché l’uso di questo sistema spazio-temporale di identificazione vada a buon fine è il nostro essere in grado non solo di identificare un particolare ma anche di reidentificarlo con successo. Reidentificare un particolare

significa considerarlo come *lo stesso particolare*, nel caso in cui la nostra percezione del particolare sia stata discontinua però, un ipotetico scettico potrebbe sostenere che tutto quello di cui disponiamo è una qualche forma di identità qualitativa e che non sia possibile affermare con certezza che i due particolari siano numericamente identici. Ancora una volta Strawson risponde facendo appello al concetto di sistema spazio-temporale sostenendo che questo ipotetico scettico in realtà sta implicitamente affermando quanto cerca di negare. Nel caso in cui non volessimo accettare all'interno del nostro schema concettuale la possibilità di una reidentificazione numerica di un qualsiasi particolare che non sia stato osservato in maniera non discontinua dovremmo possedere l'idea di un nuovo sistema spaziale per ogni nuovo continuo segmento di osservazione, stando così le cose però ogni nuovo sistema sarebbe completamente indipendente dagli altri e verrebbe quindi anche meno il dubbio riguardo a se un dato oggetto sia lo stesso in due diversi sistemi, questo infatti avrebbe senso solo se i sistemi diversi non fossero indipendenti l'uno dall'altro, ovvero se i due sistemi fossero parti di un altro che li comprende entrambi. La condizione di esistenza di questo macro-sistema però è precisamente l'esistenza di criteri attraverso i quali è possibile reidentificare come lo stesso particolare due oggetti presenti in due sotto-sistemi diversi.

Dopo aver presentato la sua teoria su come siamo in grado di identificare e reidentificare i particolari ai quali linguisticamente ci riferiamo, Strawson si chiede se tra questi particolari ci sia una categoria basilare, ovvero sul riferimento alla quale si fonda la nostra possibilità di riferirci alle altre. In questa indagine Strawson procede per gradi, eliminando prima quelli che chiama "private particulars" come gli eventi mentali e i dati di senso, in quanto sicuramente la loro identificazione dipende dall'identificazione di un altro tipo di particolari, ovvero le persone. Esclude similmente anche le entità teoriche o microscopiche della fisica in quanto necessariamente dipendenti da oggetti più grandi ed osservabili. In un primo tentativo di soluzione al problema Strawson individua i possibili candidati al ruolo di particolari basilari in ciò che chiama "material bodies", nelle persone e nei luoghi, in quanto questi sono pubblicamente osservabili e la loro identificazione non dipende dal riferimento ad altri particolari. Anche nel momento in cui prendiamo in considerazione la reidentificazione dei particolari sembra che i corpi materiali siano l'unica categoria di oggetti in grado di svolgere il ruolo di particolari basilari. Nel caso di una discontinuità nel flusso dell'osservazione infatti, la reidentificazione deve essere dipendente da una continuità dell'esistenza del particolare nello spazio all'interno di un sistema spazio-temporale condiviso, continuità garantita con successo dai corpi materiali.

Questo brevissimo riassunto del primo capitolo di *Individuals* (1959) aveva la funzione di delineare la cornice concettuale all'interno della quale si inserisce la concezione della persona formulata da Strawson. Abbiamo visto come per lui tra le diverse categorie di particolari solo i corpi materiali possono essere considerati basilari in quanto la loro identificazione e reidentificazione non dipende da nessun altro tipo di particolari. Nel secondo capitolo, che in questo contesto non verrà trattato esaustivamente, Strawson mette alla prova questa sua conclusione considerando la possibilità dell'esistenza di sistemi concettuali diversi da quello spazio-temporale al quale ha fatto riferimento nel primo capitolo. In particolare considera l'esistenza di un universo puramente auditivo cercando di verificare se in un simile sistema si mantenga la possibilità di reidentificare un suono come lo stesso suono precedentemente udito.

Strawson però arriva ad affermare che in un simile mondo non sembra esistere un metodo per distinguere tra ciò che appartiene ad una coscienza e ciò che non le appartiene, facendo così venire meno le condizioni per l'esistenza di una coscienza non solipsistica. In un universo in cui esistono solo suoni infatti l'ascoltatore dovrebbe considerarsi solo come un suono tra altri suoni ed avere la nozione di un suono che ha esperienza di altri suoni. In un simile sistema concettuale quale metodo potrebbe esistere per distinguere ciò che viene udito da chi lo ode?

Attraverso queste considerazioni Strawson arriva a chiedersi quali siano le condizioni che rendono possibile l'esistenza di una coscienza non solipsistica, cosa rende possibile la distinzione tra percepito e percepente e perché attribuiamo a noi stessi i nostri stati di coscienza? Strawson inizia la sua indagine considerando il modo in cui comunemente parliamo di noi stessi. Siamo soliti attribuirci sia azioni, intenzioni, pensieri, sensazioni, ricordi eccetera, sia caratteristiche che condividiamo con gli oggetti materiali come peso, colore, forma eccetera. Questa distinzione sembra portarsi dietro un'altra domanda che andrebbe a sommarsi alla prima ovvero: perché attribuiamo i nostri stati di coscienza alla stessa cosa alla quale attribuiamo alcune caratteristiche fisiche? Una possibile risposta ad entrambe le domande risiede nel considerare l'unicità del ruolo svolto dal nostro corpo nel condizionare la nostra percezione, per ogni persona esiste un unico corpo che occupa una posizione causale in relazione alla sua percezione sia di se stesso che di tutto ciò che esperisce. Questo fatto però per l'autore rende soltanto merito del particolare attaccamento che io provo per il mio corpo, mentre non spiega il motivo per cui io dovrei attribuire i miei stati di coscienza a qualcosa, non spiega l'uso che facciamo del termine "Io", non spiega il nostro concetto di persona.

Strawson procede nella sua indagine considerando come questa questione sia stata affrontata da due punti di vista radicalmente diversi, ovvero il dualismo cartesiano e quella che chiama la "no-ownership" theory, punto di vista che sembra essere stato sostenuto da Wittgenstein per un certo periodo e da Schlick. Entrambe le teorie rispondono a questo quesito sostenendo che i problemi appena sollevati sono semplicemente frutto di una confusione linguistica, secondo queste teorie è solo apparentemente che noi attribuiamo sia stati mentali che caratteristiche fisiche alla stessa cosa, che entrambe le categorie di predicati vengano attribuite allo stesso soggetto. Nella seconda di queste teorie, inoltre anche la prima di queste domande si basa solo su una confusione linguistica in quanto non esisterebbe alcun soggetto al quale appartenerebbero gli stati di coscienza.

Secondo il dualismo cartesiano quando ci riferiamo ad una "persona" stiamo in realtà riferendoci a due entità distinte, una *Res Extensa* ed una *Res Cogitans*, ognuna delle quali è soggetto di una categoria di predicati, di quelli di natura fisica la prima e di quelli mentali la seconda. Questa separazione, come ci viene fatto notare da Strawson, anche se risponde alla domanda sul perché attribuiamo queste due categorie diverse di predicati alla stessa cosa, "invita" invece l'altra domanda ovvero perché attribuiamo gli stati mentali ad un soggetto?

La teoria che Strawson chiama "no-ownership theory" è quella che prende spunto dall'affermazione di Lichtemberg, citata sia da Wittgenstein che da Schlick, che sostiene che invece di "Io penso" dovrei semplicemente dire "c'è un pensiero". Secondo questa teoria è a causa dell'unicità del ruolo di un particolare corpo materiale nel costringere l'esperienza sensoriale di una persona che ha luogo la confusione linguistica in base alla

quale attribuiamo un'esperienza sensoriale ad un particolare corpo materiale. L'attribuzione di un particolare stato di coscienza ad un corpo potrebbe, secondo i sostenitori di questa teoria, avere senso se continuiamo a considerare il corpo come "possessore" dello stato di coscienza nel senso in cui quest'ultimo dipende da esso e potrebbe essere dipeso da qualche altro corpo, l'errore vero e proprio si avrebbe nel momento in cui nella nostra concezione il "possessore" dello stato di coscienza viene a identificarsi con qualcosa di diverso, che chiamiamo ad esempio Ego, la cui unica funzione è rendere merito dell'attribuzione dello stato di coscienza. Nel concetto stesso di "possesso" però è compresa la possibilità che quanto viene posseduto da una cosa possa essere logicamente trasferito ad un'altra, da ciò segue che solo le cose che possono essere logicamente trasferite da un possessore ad un altro possono essere possedute. Dal momento che questo però è palesemente falso nel caso degli eventi mentali, ne segue che questi non sono genuinamente posseduti da nessuno se non unicamente nel senso dell'essere causalmente dipendenti dallo stato di un particolare corpo. Se la funzione dell'Ego era unicamente quella di rendere merito del possesso degli stati mentali in un'accezione di "possesso" che li rendeva logicamente trasferibili quindi, la stessa nozione di Ego deve essere eliminata.

Secondo Strawson questa teoria è incoerente in quanto i suoi sostenitori utilizzano lo stesso significato di "possesso" che con la loro teoria stanno cercando di negare. Nel momento in cui un sostenitore di questa teoria cerca di tradurre linguisticamente il fatto contingente che dal suo punto di vista fa insorgere l'illusione dell'Ego, necessariamente questo avrà una forma del tipo: "tutte le mie esperienze sono possedute da (unicamente dipendenti dallo stato di) un corpo B". Ogni sforzo di eliminare il "mie" o qualsiasi altra espressione che attribuisca le esperienze a qualcosa però, porta ad un'affermazione del tipo: "tutte le esperienze sono unicamente dipendenti dallo stato di un corpo B" la quale è palesemente falsa. Quello che un difensore di questa teoria sta cercando di sostenere quindi, è che "tutte le esperienze possedute da una certa persona P sono contingentemente dipendenti dallo stato di un corpo B", ma allora non può sostenere che "le esperienze di una certa persona P" significa la stessa cosa che "le esperienze contingentemente dipendenti dallo stato di un corpo B" perché altrimenti la proposizione sopra riportata non sarebbe più contingente, come richiesto dalla teoria, ma analitica. Le esperienze di cui si sta parlando devono quindi appartenere ad una certa classe di esperienze i cui membri di fatto sono contingentemente dipendenti da un corpo, ma la caratteristica che definisce questa classe di esperienze è di fatto quella di "appartenere" ad una certa persona, dove l'idea di appartenenza appena espressa è esattamente quella che i sostenitori di questa teoria stavano cercando di negare.

Quello che questa teoria non riesce a spiegare secondo Strawson è il fatto che condizione necessaria della possibilità di attribuire stati di coscienza o esperienze a noi stessi è il nostro essere in grado di attribuirle anche agli altri. Questo significa che "provare un dolore" ha lo stesso senso se applicato a noi stessi o ad altri nonostante i metodi per verificare le proposizioni "io sto provando un dolore" e "lui sta provando un dolore" sono radicalmente differenti. Secondo Strawson, l'errore commesso da Hume nel sostenere che era per lui impossibile trovare introspettivamente se stesso ma tutto ciò che era in grado di trovare erano solo percezioni, era il fatto che Hume stava cercando il concetto primitivo dell'unità di coscienza, senza tenere presente il fatto che perché si abbia questo concetto di unità è necessario che si disponga simultaneamente di

un modo per differenziare le coscienze individuali, ma questo si ha solamente se si inseriscono nel quadro generale anche le caratteristiche corporee della persona. Questo stesso problema è condiviso anche dal dualismo cartesiano, separando e rendendo ugualmente primitivi il pensiero e l'estensione infatti Cartesio ha lasciato irrisolto il problema di come questi due concetti fondamentali siano in relazione tra loro e ha reso impossibile concepire coscienze individuali differenti, altre menti. La soluzione di Strawson è quella di rendere primitivo unicamente il concetto di persona e di considerare il corpo e l'Ego come concetti secondari analizzabili all'interno di esso.

The concept of a person is logically prior to that of an individual consciousness. The concept of a person is not to be analysed as that of an animated body or of an embodied anima. This is not to say that the concept of a pure individual consciousness might not have a logically secondary existence, if one thinks, or finds, it desirable. We speak of a dead person - a body - and in the same secondary way we might at least think of a disembodied person. A person is not an embodied ego, but an ego might be a disembodied person, retaining the logical benefit of individuality from having been a person.
(Individuals. Cap 3, 103)

Per Strawson non possiamo avere l'idea di noi stessi senza essere simultaneamente in grado di identificare le altre persone come altre coscienze individuali. Per chiarire meglio quanto sta sostenendo Strawson affronta anche il problema da un punto di vista linguistico distinguendo tra M-predicati, ovvero predicati che si applicano sia alle persone che ai corpi materiali, e P-predicati, ovvero predicati che si applicano unicamente alle persone. La tesi sostenuta in quest'ottica sarà quindi che imparare ad attribuire i P-predicati significhi imparare simultaneamente come applicarli a se stessi ed agli altri, nonostante i metodi di verifica dell'applicazione sembrano essere radicalmente differenti.

A prima vista sembra che l'applicazione dei P-predicati agli altri sia un procedimento necessariamente legato all'osservazione del loro comportamento, mentre l'applicazione di questi a se stessi non lo è, secondo Strawson non esiste un metodo logicamente primario per apprendere ad applicare i P-predicati ad esempio agli altri sulla base dell'osservazione, dal quale deriva un metodo logicamente secondario per apprendere ad attribuirli a noi stessi. Apprendere ad utilizzare i P-predicati significa simultaneamente apprendere ad attribuirli sia a noi stessi sia agli altri. Per poter capire questo tipo di concetti una persona deve essere in grado di utilizzarli in entrambi i modi. Nel momento in cui noi distinguiamo tra un comportamento depresso e la sensazione della depressione siamo pericolosamente in bilico tra comportamentismo e scetticismo, la soluzione a questo problema è per Strawson capire che per far sì che si possa dire che X è depresso, necessariamente il concetto di depressione deve coprire sia la depressione provata ma non osservata, sia la depressione osservata ma non provata. Se infatti consideriamo come primario l'uso auto-attributivo dei P-predicati, si apre un divario logico tra i criteri secondo i quali definiamo gli altri come ad esempio depressi, e lo stato dell'essere depressi, in questo modo infatti un comportamento depresso non è altro che un segno di depressione, ma questo può essere diventato solo attraverso il nostro aver osservativamente trovato una correlazione tra la depressione e quel tipo di comportamento. Se al contrario poniamo come primaria l'attribuzione di questi

predicati agli altri arriveremo a pensare che tutto ciò che costituisce il significato di questi predicati sono i loro criteri di attribuzione. Questo però significa dimenticare l'uso auto-attributivo di questi predicati che non si basa sull'osservazione di un comportamento. Questi predicati non hanno due significati diversi, ma uno stesso che comprende i due casi di attribuzione.

L'errore, in entrambi i casi è rappresentato dal tracciare una linea di demarcazione tra se stessi e gli altri, mentre nel momento in cui ad esempio stiamo partecipando ad un gioco di squadra o un'altra azione che coinvolge sia movimenti del corpo, sia intenzioni e stati di coscienza, non esitiamo a considerare le nostre azioni e quelle degli altri come dello stesso tipo ed a capirle per il loro essere guidate da intenzioni simili e condivise grazie al loro comune appartenere a persone.

Essere una persona quindi secondo Strawson significa possedere e poter attribuire a se stessi ed alle altre persone sia P-predicati che M-predicati. Il concetto di persona che viene plasmato nelle pagine di *Individuals* è un concetto primitivo e basilare in base al quale secondariamente possiamo concepire concetti diversi come ad esempio quello di un ego privato del suo corpo o viceversa. In quest'ottica quello di persona diventa un concetto fondamentale della nostra cognizione che non può essere ridotto ad altri più basilari come avveniva ad esempio per il dualismo cartesiano. Conseguentemente a questa formulazione non potranno esistere criteri per l'identificazione di una persona come la stessa in due tempi diversi che siano esclusivamente psicologici, come avveniva nella teoria di Locke, o fisici come viene sostenuto dalle teorie fisicaliste sull'identità personale, ma similamente a quanto avviene nel caso dell'attribuzione dei P-predicati e degli M-predicati, entrambi gli aspetti saranno simultaneamente presenti ed inscindibili nella nostra concezione di identità personale, il concetto stesso di identità personale comprenderà entrambe le parti senza poter essere ridotto a nessuna delle due, mentre sarà invece possibile il contrario, ovvero concepire la continuità psicologica o fisica come parti volutamente separate della fondamentale continuità della persona.

A.J.Ayer (1963) Ha obiettato a Strawson che ponendo il concetto di persona come primitivo e rifiutando di considerare sia il significato relativo all'attribuzione ad altri dei P-predicati, sia il senso auto-attributivo come fondamentale

"Mr Strawson hopes to secure the advantages, and at the same time avoid the difficulties, of both these lines of approach. He admits that the basis on which we ascribe experiences to others is different from that on which we ascribe them to ourselves, but he denies that, in the case of other people, we are reduced to making an inductive inference. "

(1963. 91).

Il problema secondo Ayer è che veniamo lasciati da questa formulazione senza un tentativo di soluzione al problema. Strawson infatti dopo aver eliminato le due possibili linee di approccio, non cerca di spiegarci come questo tipo di predicati sia possibile. Viene inoltre mantenuto da Strawson l'assunto che la nostra osservazione del comportamento degli altri costituisca un criterio logicamente adeguato per la nostra attribuzione di esperienze alle altre persone, nonostante rifiuti che ciò avvenga per analogia con la nostra auto-attribuzione di quelle esperienze. Il punto secondo Ayer è che Strawson cerca di definire il concetto di logicamente adeguato prendendo ad

esempio un gioco di carte in cui un segno esteriore costituisce un criterio logicamente adeguato perché una carta venga chiamata ad esempio regina di cuori, mentre il predicato “essere la regina di cuori” prende il suo significato dall’intera struttura del gioco. Allo stesso modo per Strawson, i criteri fisici e comportamentali pur rimanendo logicamente adeguati per la nostra attribuzione a persone diverse da noi di esperienze che presuppongono una coscienza interna, non rappresentano l’intero significato di questi predicati. Secondo Ayer però nonostante sia indubitabile che il predicato “essere la regina di cuori” prenda il suo significato anche in base alle funzioni di quella carta all’interno della struttura del gioco, queste rimangono tali sulla base di una convenzione che le ha definite e da cui ne traiamo il significato. La nostra attribuzione di dolore ad una persona dall’aspetto sofferente invece non dipende da nessun tipo di convenzione, non ci sarebbe alcuna contraddizione logica infatti, nel vedere un viso sofferente e continuare a negare che quella persona stia provando dolore. Sembra insomma, secondo Ayer, che Strawson rifiutando di servirsi dell’argomento dell’analogia, non riesca più a conciliare i due significati diversi dei P-predicati ed a garantire un criterio logicamente adeguato per il loro uso.

2.2 L’approccio fisicalista al problema dell’identità personale.

L’approccio criterialista al problema dell’identità personale si pone in netto contrasto con la visione di Strawson fin nelle sue fondamenta teoriche. Quanto sostenuto dai criterialisti è infatti la tesi che l’identità personale possa essere definita attraverso criteri necessari e sufficienti, ovvero esattamente quanto viene negato dalla teoria di Strawson, in modo tale che la risposta alla domanda: “Che cosa rende una persona la stessa in due tempi differenti?” possa coerentemente avere la forma: “ x ad un tempo t ’ è la stessa persona di y ad un tempo t ” *se e solo se...*”. I possibili candidati a completare questa formula si dividono in criteri fisici e criteri psicologici. La presentazione delle due linee di pensiero corrispondenti alle due categorie di criteri sarà oggetto delle restanti parti di questo capitolo.

Per quanto riguarda i criteri di natura fisica la teoria dell’identità personale che si presenta come la più semplice e naturale è quella che utilizza come criterio necessario e sufficiente l’identità del corpo. Secondo questa teoria una persona viene considerata la stessa in due tempi differenti se possiede lo stesso corpo. In questa accezione l’identità personale non è diversa dall’identità di qualsiasi altro artefatto in generale. La persistenza di un artefatto attraverso il tempo infatti, non consiste necessariamente nella persistenza della stessa materia che lo costituisce. Come esemplificato dall’ormai celebre esempio della nave di Teseo infatti, un artefatto può venire riparato e le sue parti possono venire progressivamente sostituite fino a renderlo costituito interamente di materia differente lasciando però intatta l’identità di esso. La stessa cosa avviene per gli organismi viventi, come ad esempio un albero o appunto il corpo umano, in cui in un periodo di tempo abbastanza lungo la totalità delle sue cellule, escluse quelle del cervello, viene gradualmente sostituita. Quanto viene richiesto da questa teoria perché

venga preservata l'identità personale è che il corpo che costituiva la persona al tempo t' abbia subito un rinnovamento più o meno graduale della sua materia, pensiamo ad esempio ad un trapianto di qualche organo, in modo tale da poterlo definire come lo stesso corpo al tempo t'' .

Nonostante questo criterio sembri avvicinarsi molto alla nostra esperienza di tutti i giorni questa teoria è stata in larga parte abbandonata a causa dell'esistenza di alcune situazioni possibili in cui il nostro concetto di identità personale sembra portarci in direzioni diverse. Mi riferisco in particolare all'esperimento mentale formulato da Shoemaker (1963) in cui si immagina che in un futuro più o meno lontano l'uomo sia in grado di eseguire a tutti gli effetti un trapianto di cervello. Immaginiamo in questo contesto che uno scienziato trapianti il cervello di un uomo Mr. Brown, nel cranio di un altro uomo, Mr. Robinson. La persona risultante dal trapianto, che chiameremo Brownson, è una persona in piena salute, in grado di comunicare ed interagire con gli altri, con il corpo di Mr. Robinson, ma indistinguibile nei ricordi nella personalità e nei tratti caratteriali da Mr. Brown a causa del possesso del suo cervello. Se supponiamo inoltre che i corpi di Mr. Brown e di Mr. Robinson fossero esattamente simili diventa difficile negare che la persona risultante dal trapianto non sia identica a Mr. Brown. Queste considerazioni hanno spinto diversi filosofi a sostituire il criterio dell'identità del corpo con quello di una sua parte considerata fondamentale, ovvero il cervello, di modo tale che l'identità personale si basi sempre nella persistenza di una entità di tipo fisico. Secondo questa visione del problema quindi, il criterio necessario e sufficiente per determinare l'identità personale è l'identità del cervello.

Ad ogni modo questa modificazione della teoria che stabiliva l'identità del corpo come criterio necessario e sufficiente per l'identità personale non sembra essere abbastanza radicale per far fronte alla totalità dei casi possibili. Così come è stato facilmente immaginato un caso in cui l'identità del corpo non risultasse essere un criterio adeguato per l'identità personale, sembra infatti altrettanto facile immaginarne altri, ugualmente scientificamente comprensibili, in cui non lo è neanche l'identità del cervello. Consideriamo ad esempio il fatto che il nostro cervello è composto da due emisferi distinti e tra loro collegati attraverso il corpus callosum, ovvero molto semplicisticamente un intreccio di fibre. Ognuno dei due emisferi tipicamente svolge compiti diversi dall'altro con il quale si divide il lavoro, in un adulto ad esempio saranno tipicamente molto più sviluppate le capacità matematiche e linguistiche nell'emisfero sinistro, mentre in quello stesso emisfero la musicalità ad esempio rimarrà al livello di quella di un bambino e viceversa nell'emisfero destro. Può succedere che in determinati casi conseguenti ad una malattia o ad un incidente ad una persona rimanga un solo emisfero funzionante, costringendo l'unico emisfero rimasto a svolgere tutto il lavoro da solo ed a sviluppare quelle funzioni che erano rimaste in lui ad un livello elementare. È stato notato attraverso lo studio di casi come questi che un unico emisfero è sufficiente alla sopravvivenza di una persona e che l'unico emisfero rimasto diventa rapidamente in grado di svolgere anche quei compiti che precedentemente erano prevalentemente svolti dall'altro. Consideriamo a questo punto un caso simile a quello sopra riportato in cui però solo un emisfero del cervello di Brown viene trapiantato nel corpo di Robinson. Coerentemente con quanto affermato in precedenza sembra che anche in questo caso venga preservata l'identità personale del donatore del cervello, o meglio di una parte di esso. Queste considerazioni hanno portato alcuni filosofi come

Wiggins (1967, 2001) a proporre come criterio di identità personale quello che viene chiamato il Physical Criterion, ciò che viene in quest'ottica considerato un criterio sufficiente per l'identità personale non è il cervello, bensì soltanto una parte sufficiente di esso perché sia mantenuta una continuità psicologica.

Anche quest'ultima e meno restrittiva formulazione però non è esente da problemi, Bernard Williams (1970, 1973) ad esempio immagina un caso, che rasenta la fantascienza, in cui la totalità delle conoscenze, dei ricordi e in generale del contenuto del cervello di un uomo, viene trasferita all'interno di un dispositivo di immagazzinamento, per poi essere ritrasferita all'interno di uno stesso o di un altro cervello vuoto. Secondo Williams è chiaro che l'uomo, al quale supponiamo sia stato sostituito il cervello a causa di un danneggiamento, debba essere considerato la stessa persona, ma in questo caso allora anche il criterio fisico di identità deve essere abbandonato. Quello che Williams ha sottolineato attraverso il suo caso fantascientifico è che nelle ultime due teorie fisicaliste presentate il cervello svolgeva un ruolo fondamentale ed era considerato l'unica parte del corpo realmente rilevante a causa del suo svolgere la funzione di mantenimento della continuità psicologica della persona, ma nel momento in cui questa funzione non viene più svolta dal cervello l'identità di questo perde la sua caratteristica di condizione necessaria per l'identità personale.

Attraverso una successiva modificazione del suo esperimento mentale Williams inoltre attacca anche il proporsi come condizione sufficiente per l'identità personale del criterio fisico. Supponiamo che prima che il cervello di Brown venga trapiantato all'interno del corpo di Robinson questo venga ripulito di tutti i suoi contenuti e riempito nuovamente, sempre per mezzo del dispositivo di immagazzinamento, di tutti i contenuti del cervello di Robinson. A questo punto dopo il trapianto Brownson non solo ha il corpo di Robinson, ma ha anche tutti i ricordi ed i tratti caratteriali di Robinson. Se nella prima formulazione di questa storia eravamo arrivati ad affermare che Brownson era la stessa persona di Brown, difficilmente in questa versione potremmo negare che Brownson è la stessa persona di Robinson, ma stando così le cose sembra che il criterio fisico di identità debba essere definitivamente abbandonato.

Un'ultima obiezione al criterio fisico di identità risiede nel fatto che quando in contesto filosofico si utilizza il termine "persona" questo non viene considerato un sinonimo di "essere umano", ma piuttosto viene utilizzato nell'accezione che ne dava Locke con la sua definizione di "*a thinking intelligent being with reason and reflection that can consider itself as itself, the same thinking being, in different times and places.*" Persona in questa accezione è quindi un termine funzionale che si applica ad esseri dotati di determinate capacità. È quindi possibile in linea di principio che esistano nell'universo persone non umane nelle quali non esiste un organo le cui funzioni siano paragonabili a quelle svolte dal cervello umano ed alle quali quindi il criterio fisico di identità risulterebbe inapplicabile.

Una teoria, sempre di matrice fisicalista, molto più radicale di quelle prese in considerazione fino ad ora è quella sviluppata da Eric Olson (1997, 1999), che definisce la sua visione come l'approccio biologico al tema dell'identità personale. Caratteristica fondamentale della sua tesi è il netto rifiuto del fatto che qualsiasi tipo di continuità psicologica possa avere la funzione di criterio necessario o sufficiente per l'identità personale, indipendentemente dal fatto che questa sia garantita da una parte fisica del nostro organismo o meno. Secondo Olson l'unico criterio necessario e sufficiente al

nostro persistere è la continuità biologica, ciascuno di noi infatti, sopravvive solo nel caso in cui continuano le sue funzioni animali. Alla base della teoria di Olson c'è una riformulazione del problema dell'identità personale. Più che essere interessato ai criteri sui quali si fonda il nostro persistere come le stesse persone infatti, Olson, dopo aver sostenuto che la categoria che meglio ci definisce non è quella di persone, ma di esseri umani, si chiede semplicemente in base ai quali criteri noi possiamo persistere come gli stessi. Nell'interpretazione di Olson il predicato "essere una persona" non rappresenta più un predicato di tipo sortale o sostanziale ma solamente un predicato di fase legato al nostro possesso di determinate capacità e funzioni, in quest'ottica quindi "essere una persona" non rappresenta qualcosa che definisce ciò che fondamentalmente siamo ma solo l'attribuzione di caratteristiche transitorie a qualcosa di più fondamentale, ovvero il nostro essere esseri umani. In questo senso nell'arco della vita di un uomo è possibile che questo smetta di essere una persona senza cessare di essere un essere umano, come ad esempio nel caso in cui una persona a causa di una malattia o di un incidente si trovi in uno stato vegetativo. Nonostante l'interpretazione che Olson ci fornisce dell'essere una persona non sia lontana da quelle psicologistiche, ciò che lui sostiene è il fatto che "persona" non è il termine che ci caratterizza fondamentalmente affermando che è la sola continuità biologica che garantisce l'identità di un essere umano in due tempi diversi.

Molte obiezioni sono state dirette ad Olson non tanto per il fatto che secondo questa teoria una persona può esistere solo per una fase della vita dell'organismo in modo tale che la sua persistenza sia da esso logicamente distinta, ma più propriamente per il fatto che nel momento in cui io mi riferisco a me stesso attraverso il termine "io", ciò a cui mi sto riferendo non è il mio organismo animale ma la mia esistenza come persona, in modo tale che nel momento in cui cerchiamo di definire cosa significhi persistere come noi stessi attraverso il tempo è al nostro persistere come le stesse persone che ci stiamo riferendo. Nonostante la formulazione di Olson abbia il difetto di ridurre il problema della persistenza della nostra identità ad un "affare da biologi" incapace di tenere conto degli aspetti e delle nostre caratteristiche distintive in base alle quali siamo soliti definirci, questa formulazione ha il merito di aver messo in luce il fatto che l'insistenza sui soli tratti psicologici tipica di molte teorie criterialiste ha impoverito la categoria di persona fino a renderla quasi una categoria funzionale non più in grado di mantenere il ruolo di concetto sortale che anche Locke le attribuiva. Il punto di rottura tra Olson e le teorie psicologistiche moderne risiede nel suo rifiuto, nel momento in cui si accetta il fatto che il predicato "essere persona" si applichi solo ad una certa fase del nostro essere umani caratterizzata dal possesso di alcune caratteristiche distintive, del fatto che questo predicato identifichi comunque una sorta di entità ulteriore rispetto all'essere esseri umani. Olson nega il fatto che nel momento in cui apparteniamo alla categoria "persone" noi siamo essenzialmente persone ed il nostro permanere come identici si leghi quindi al nostro permanere come le stesse persone. Secondo Olson io posso permanere identico anche nel momento in cui cessi di essere una persona essendo io primariamente un animale umano i cui criteri di identità sono esclusivamente biologici e solo secondariamente e fatalmente una persona.

2.3 La “standard view” e i criteri di natura psicologica.

La corrente che ha avuto più seguito nell'ultimo secolo e che viene infatti definita “standard”, è quella del criterialismo psicologico. Sotto questo nome troviamo quelle teorie che si propongono di affermare la necessità di un criterio psicologico-mnemonico per l'identità personale, liberandolo però dai problemi e dalle incongruenze tipiche della originaria formulazione lockeana. Contrariamente a quanto avveniva nel caso di Locke, i moderni sostenitori di questa visione dell'identità personale sono liberi dalle costrizioni di tipo teologico ed escatologico che invece costringevano ai propri fini la teoria lockeana, di modo tale che non di rado nel novecento il criterialismo psicologico si è definito materialista, non intendendo assolutamente negare che le caratteristiche di continuità psicologica definite necessarie all'identità fossero legate ad una parte fisica del nostro corpo, come è implicito nel caso di Brownson di Sydney Shoemaker (1963). Nel difendere una visione che trova nella continuità psicologica il suo criterio necessario e sufficiente molti criterialisti moderni infatti, non intendono assolutamente affermare l'indipendenza della coscienza da un corpo o inserirsi nel dibattito sulla relazione mente-corpo, ma solamente individuare una funzione psicologica e difenderla come definitoria del concetto di identità personale, rendendo la nostra sopravvivenza indipendente da quella di un corpo particolare.

I problemi che i sostenitori di una qualche forma di criterialismo psicologico si trovano costretti ad affrontare sono grossomodo gli stessi che abbiamo trovato formulati nelle obiezioni alla teoria di Locke considerate nel primo capitolo. In particolare l'obiezione di circolarità e di non transitività, esemplificata da Reid nella storia del “Galland Officer”, costringeranno molti psicologi moderni ad adattare le loro formulazioni in base ad esse. La principale preoccupazione sia di Shoemaker che di Parfit nel loro definire un concetto di quasi-memoria sul quale fondare la loro teoria è infatti quella di evitare l'accusa di circolarità, la quale si è rivelata come uno degli ostacoli più difficili da superare.

Nella prima formulazione della sua teoria presentata in “*Self-Knowledge and Self-Identity*” (1963) Shoemaker attraverso l'esperimento mentale di Brownson si proponeva semplicemente di mostrare che nella nostra concezione di identità personale ci fosse almeno una componente di matrice psicologica accanto ad una materiale, successivamente ha però adottato una visione più radicalmente psicologista cercando di superare i paradossi tipici delle teorie che attribuiscono alla continuità psicologica il ruolo di criterio necessario e sufficiente per la persistenza di una stessa persona. I principali ostacoli che Shoemaker in “*Persons and Their Pasts*” (1970) e “*Personal Identity*” (1984) ha cercato di superare sono quello di trovare una giustificazione alla credenza della generale veridicità delle memorie su se stessi e soprattutto quello di superare l'obiezione di circolarità considerata nel primo capitolo, definendo una relazione tra il protagonista dei ricordi di una persona e quella stessa persona che non presupponga di per sé l'identità personale. A questo scopo è finalizzata la sua introduzione del concetto di “quasi-memory knowledge”, il quale definisce un “quasi-ricordare” che si differenzia dal normale “ricordare” per il suo non garantire che la cosa ricordata appartenga alla stessa persona che la ricorda. In quest'ottica la le garanzie che

la mia credenza che ciò che quasi-ricordo sia un avvenimento del mio passato deriverebbe dalla sua congruenza con altri quasi-ricordi, dal fatto che la falsità del ricordo, in ragione dei suoi contenuti spazio-temporali, implicherebbe reduplicazioni implausibili e dall'appropriatezza della relazione causale che concatena i miei quasi-ricordi fino ad arrivare al mio stato presente. Secondo questa definizione quindi io posso dire di quasi-ricordare una passata esperienza o quant'altro nel caso in cui:

- (1) Mi trovo attualmente in uno stato *come* quello del ricordare qualcosa.
- (2) Il contenuto di questo stato è congruente con la natura dell'esperienza passata.
- (3) Questo stato è collegato in maniera appropriata al mio stato di coscienza passato relativo all'esperienza quasi-ricordata attraverso una catena causale di quasi ricordi.

Shoemaker chiama questa catena causale che collega in maniera veridica una azione o esperienza passata ad un quasi-ricordo presente una "M-type causal chain". Con questa formulazione Shoemaker ha esternalizzato in un impianto olistico le condizioni di affidabilità dei miei quasi-ricordi rendendo perciò una persona incapace di fondare la sua credenza nella veridicità di un ricordo unicamente sul modo che questo ha di presentarsi alla sua coscienza.

Questa esternalizzazione delle garanzie della veridicità di un ricordo ha sollevato molti dubbi tra i filosofi, in particolare Gareth Evans (1982) ha sottolineato il fatto che pensare alla memoria come ad una forma di conoscenza neutrale dal punto di vista dell'identità sia decisamente controintuitivo. Quello che Evans intende affermare con questa obiezione non è il fatto che la nostra memoria non possa essere fallibile, ma che la fallibilità della nostra memoria non risiede nella possibilità di un errore nell'identificazione del protagonista del ricordo, ma nella possibilità di un errore nell'attribuzione di determinate proprietà e caratteristiche al protagonista del ricordo che rimane per definizione identico a chi possiede il ricordo.

Attraverso la nozione di quasi-memory Shoemaker è riuscito quindi ad evitare l'obiezione di circolarità, al prezzo però dell'autonomia e dell'indipendenza del criterio di identità che ha ora bisogno anche di elementi esterni al mero flusso di coscienza per essere verificato. Secondo Shoemaker nel momento in cui io mi chiedo se sia io il protagonista di una passata azione o esperienza di cui ho una quasi-memoria, non cerco di dare una risposta a questa domanda facendo appello ad un criterio di identità personale, cosa che mi farebbe cadere nell'obiezione di circolarità, al contrario la domanda che mi sto ponendo è se ogni nodo della catena causale di quasi ricordi che mi collega a quell'esperienza o azione sia collegato appropriatamente al suo vicino. Nel caso in cui io riesca a verificare che ciò sia avvenuto, avrò stabilito un criterio sufficiente di identità personale.

Nell'arco del novecento però sono state dirette molte obiezioni proprio verso questa pretesa del criterio psicologico di essere un criterio sufficiente. Una delle più efficaci è l'argomento della duplicazione prodotto da Bernard Williams (1973). In questo argomento Williams suppone che una persona B a seguito di una serie di cambiamenti psicologici mostri una perfetta continuità psicologica e mnemonica con una certa persona precedentemente deceduta A. Secondo le teorie che affermano che il criterio psicologico sia un criterio sufficiente per stabilire l'identità di una persona ne

seguirebbe che A e B siano quindi la stessa persona. Se però un tale cambiamento psicologico è avvenuto in B non possiamo logicamente escludere che tale cambiamento possa aver luogo anche in una terza persona C. A questo punto però non è più possibile sostenere che A e B siano la stessa persona perché altrimenti si avrebbe anche che B e C sono identiche, cosa che significherebbe o sostenere che la stessa persona può occupare due spazi diversi nello stesso tempo oppure ammettere il fatto che possa essere violata la transitività dell'identità personale in quanto B e C sarebbero entrambi identici ad A senza esserlo tra di loro. Un'altra obiezione che Williams riporta allo scopo di sostenere l'esistenza di uno stretto legame tra la nozione di identità e quella di continuità spazio temporale, legame che il semplice ricorso alla continuità mnemonica sembra negare, è quello che chiama "L'argomento delle conclusioni contrarie". In questo argomento Williams presenta da due punti di vista differenti un esperimento mentale in modo da condurci a conclusioni tra loro opposte. L'argomento muove dalla considerazione di un caso molto simile all'esperimento mentale di Brownson in cui ad una persona A e ad una persona B attraverso un intervento chirurgico vengono scambiati i corpi in modo tale che alla fine dell'intervento la psicologia di A risieda nel corpo di B e viceversa. Williams successivamente suppone che dopo l'intervento una delle due persone debba subire una tortura e che prima dell'intervento sia ad A che a B separatamente venga chiesto, su basi puramente egoistiche, quale corpo preferirebbero che venisse torturato. Intuitivamente sembrerebbe razionale che ciascuna delle due persone candidi alla tortura il proprio corpo di partenza identificando il se stesso post-operazione con la propria psicologia e non con il proprio corpo. Supponiamo invece che qualcuno mi comunichi che domani io sarò torturato, ma che aggiunga anche che nel momento in cui io sarò torturato tutte le mie caratteristiche psicologiche ed i miei ricordi saranno stati alterati fino a rendere la mia identità psicologica del tutto diversa. In questo caso sembra altrettanto intuitivamente che io non possa trarre da questo cambiamento nessun sollievo dal timore della tortura futura che continuo a considerare come rivolta a me stesso, lo stesso sostiene Williams accade anche nel caso in cui questa mia continuità psicologica venga scambiata con quella di un'altra persona, nonostante la nostra intuizione da senso comune ci abbia spinto a pensare diversamente, anche in quel caso è in realtà la continuità corporea ad essere decisiva per l'identità personale.

Interessante e rilevante per gli scopi di questo testo è anche la teoria riduzionista di Derek Parfit (1984). La visione di Parfit si qualifica come riduzionista in quanto sostiene che l'identità personale sia un fenomeno composito e quindi la sua analisi è riducibile ai fatti più fondamentali, secondo Parfit di natura psicologica, di cui è composto. Altra particolarità della teoria di Parfit è il suo sostenere, in certo senso in maniera analoga a quanto difeso da Olson, che ciò che realmente conta per la nostra sopravvivenza non è l'identità.

L'analisi di Parfit muove i suoi primi passi dal tentativo di migliorare il criterio di identità formulato da Locke facendo riferimento, similamente a quanto sostenuto anche da Leibniz, anche alla continuità psicologica oltre che alla sola connessione psicologica con l'azione o l'esperienza ricordata. Per Parfit si ha una continuità psicologica nel momento in cui, normalmente, una persona nell'arco degli anni sviluppi di giorno in giorno una connessione mnemonica diretta con la maggior parte degli avvenimenti del giorno prima. La continuità mnemonica così formata sussisterà anche nel momento in cui io in un tempo t''' non abbia più collegamenti mnemonici diretti con un tempo t' , ma

abbia a suo tempo stabilito una connessione diretta tra t' e t'' , e tra t'' e t''' che supponiamo essere il tempo presente. Condizione necessaria per il crearsi della continuità psicologica è quindi lo stabilirsi di giorno in giorno della connessione mnemonica con gli avvenimenti del giorno precedente, in quest'ottica quindi sia la continuità psicologica che la connessione mnemonica diretta sulla quale la prima è fondata saranno criteri sufficienti per l'identità personale.

La continuità psicologica e la connessione diretta costituiscono quella che Parfit definisce come la relazione R e che utilizza per dare una prima definizione riduzionista dell'identità personale. Nello specifico A e B sono la stessa persona in due tempi diversi se:

- (1) Tra A e B vige la relazione R.
- (2) La relazione R è data dal giusto tipo di causa.
- (3) La relazione R non ha forma ramificata.

La terza condizione esprime la necessità che la relazione R sia unica, questo requisito viene posto da Parfit allo scopo di evitare le conseguenze di situazioni controverse come quelle in cui avviene una fissione, ovvero come nel caso della separazione dei due emisferi del cervello che abbiamo già avuto modo di incontrare, o quella di un ipotetico teletrasporto che ricrei cellula per cellula una persona in un luogo diverso senza distruggere l'originale. In casi come questi si avrebbero due istanziazioni ugualmente valide della relazione R, e di conseguenza due candidati ugualmente validi per l'identità personale. Notiamo però che se in casi come questi una metà del cervello andasse distrutta o il teletrasporto distruggesse anche l'originale lasciando intatta la replica verrebbero mantenute tutte le condizioni per l'identità personale. In questo modo però si accetterebbero come rilevanti alla questione dell'identità di due oggetti x e y , anche elementi indipendenti da x e y , sulle conseguenze che questo ha sul concetto stesso di identità comunque tornerò più esaurientemente nel prossimo capitolo.

L'intento di Parfit nel definire la relazione R comunque non è tanto quello di capire quali restrizioni vanno applicate ad essa per farne un criterio di identità, ma piuttosto quello di mostrare che l'identità personale non ha per noi l'importanza che credevamo avesse, mentre l'unica cosa che conta ai fini della nostra sopravvivenza è la relazione R che sia ramificata o meno. Secondo Parfit infatti, il fatto che oltre ad esistere la relazione R abbia anche forma non ramificata, garantendo così l'identità personale, ha soltanto una rilevanza di tipo logico ma non esistenziale, ciò che conta per la mia sopravvivenza è esclusivamente la relazione R. In quest'ottica piuttosto che come persone, dovremmo descriverci come una sequenza di io successivi più o meno legati tra loro in base alla relazione R.

Il problema che Parfit deve affrontare a questo punto è sempre lo stesso, ovvero dare una definizione della relazione R che non presupponga il concetto di identità personale e per questo scopo si serve di una nozione di quasi-ricordo modificata secondo la quale io posseggo un quasi-ricordo di un'esperienza passata se:

- (1) Mi sembra di ricordare di aver avuto un'esperienza.
- (2) Qualcuno ha avuto un'esperienza.
- (3) Il mio apparente ricordo è causalmente dipendente da quell'esperienza e lo è nel

modo corretto.

Anche secondo questa definizione un quasi-ricordo risulta essere un ricordo che io potrei avere in prima persona, senza però che questo faccia riferimento ad un'esperienza del mio passato, condizione necessaria perché l'identità non venga presupposta e perché Parfit possa definire la relazione R nei termini della quasi-memoria. Definendo la relazione R in questi termini non è nemmeno necessario porre alcuna restrizione riguardo a quali tipi di catene causali siano da considerarsi giuste o meno, Parfit infatti ritiene che non sia importante stabilire in ragione di quali categorie di catene causali, siano esse affidabili o inaffidabili, si abbia la relazione R, che questa venga mantenuta attraverso il teletrasporto, la continuità del cervello o quant'altro infatti non è rilevante per la mia sopravvivenza, l'unica cosa rilevante è il fatto che venga mantenuta.

A sostegno della sua visione riduzionistica dell'identità personale Parfit propone un esperimento mentale da lui definito come solo tecnicamente e non concettualmente impossibile, quello degli spettri combinati. In questo esperimento mentale Parfit immagina la possibilità di un processo che, date due persone di partenza, attraverso modificazioni graduali e microscopiche, sostituisca le caratteristiche fisiche e psichiche dei soggetti di partenza le une con le altre fino a scambiarle completamente. In questo modo ciascuna delle due persone risulterà durante il processo avere una connessione psicologica sempre più debole con chi era in partenza e sempre più stretta con l'altra persona. Questo esperimento mentale rappresenta un forte argomento a favore della concezione graduata e riduzionista dell'identità personale portata avanti da Parfit. In questa situazione infatti, non sembra esserci una linea di demarcazione netta superata la quale avviene uno scambio di identità tra i due soggetti, non sembra esistere di fatto una "identità della persona" intesa come fatto ulteriore e non riducibile a componenti fisiche o psicologiche. In questo esperimento mentale troviamo avvalorata anche la tesi di Parfit secondo cui l'identità personale possa in certi casi rimanere indeterminata, nel senso che nonostante noi siamo a conoscenza di tutti i fatti che riguardano i suoi costituenti, possiamo comunque non essere in grado di stabilire se sussista l'identità personale o meno. Attraverso questa analisi Parfit arriva a sostenere l'irrelevanza dell'identità personale rispetto alla relazione R da un punto di vista riduzionista. Afferma infatti, che se si accetta una prospettiva riduzionista dell'identità personale non è più necessario tentare di formularne una definizione o definirne i criteri perché il concetto stesso di identità personale in questa prospettiva diventa irrilevante nel definire i termini del nostro perdurare nel tempo.

Come viene fatto notare da Carlo Gabbani (2007) una simile prospettiva sembra però voler attribuire un ruolo di maggior importanza alla relazione R rispetto all'identità personale, senza prima però aver definito e ridotto quest'ultima. Per esemplificare le conseguenze della sua obiezione Gabbani ipotizza una situazione in cui un soldato viene ferito alla testa in una battaglia e, al momento del suo risveglio dopo il trauma, non ricorda più niente del suo passato. In questo caso stando alla teoria di Parfit, che definisce l'identità personale come una relazione R non ramificata, l'uomo che si risveglia in convalescenza è un'altra persona rispetto al soldato ferito in battaglia, essendo venuta a crearsi una cesura nella sua continuità psicologica. Supponiamo però che quest'uomo dopo il risveglio rimanga sempre normalmente cosciente e che riesca a stabilire giorno per giorno una connessione mnemonica diretta con gli avvenimenti del

giorno prima fino al momento in cui viene dimesso, in modo tale che la persona che viene dimessa un mese dopo è la stessa persona che si è risvegliata dopo il trauma. Nei mesi successivi alla sua uscita dall'ospedale però, l'uomo lentamente inizia di nuovo a ricordare fatti e esperienze del suo passato fino a ristabilire completamente la continuità psicologica con la persona che era prima della battaglia in cui è stato ferito. A questo punto, se l'identità personale viene definita come una relazione R non ramificata, ci troviamo in una situazione in cui la persona che ha ritrovato la sua memoria è la stessa persona che è stata dimessa dall'ospedale, che si è risvegliata dal trauma e che è stata ferita alla testa ma la persona che è stata ferita è una persona diversa rispetto a quella che si è risvegliata in ospedale. Saremmo dunque di fronte ad una definizione di identità personale che non ne salvaguarda la transitività, proprietà logica considerata esplicitamente da Parfit stesso come irrinunciabile. Considerazioni come queste e come quelle fatte da Bernard Williams che ho descritto sopra, hanno portato Parfit a riconsiderare la plausibilità del suo criterio di identità personale fino al punto di sostenere l'irrilevanza del cercare di definire un concetto come quello di identità personale che potrebbe addirittura essere malformato.

In questo modo però Parfit arriva a conclusioni che sembrano mettere in discussione l'impianto stesso della sua teoria, nonostante arrivi a considerare come irrilevante l'identità personale e arrivi a negare l'esistenza di criteri che la possano definire a causa della sua vaghezza e malformazione, ciò nonostante introduce alla sua relazione R la condizione del non aver forma ramificata al solo scopo di renderla un criterio adeguato di identità personale. Il motivo stesso per il quale inizialmente Parfit rifiuta il criterio di Locke è che lo considera un criterio inadeguato per la riduzione dell'identità personale per gli stessi motivi per i quali si è rivelato inadeguato anche il suo.

Parfit insiste frequentemente nel presentare le sue tesi che "ciò che conta" non è l'identità personale ma la relazione R. Sembra però possibile immaginare sia casi in cui la relazione R non abbia grande importanza, sia casi in cui l'assenza della relazione R con un'esperienza del mio passato non cancelli l'importanza di questa in quanto parte della mia storia personale. Gabbani (2007) immagina a questo proposito il caso di una persona che non riesca a incamerare ricordi per più di ventiquattro ore, in modo tale che ogni mattina riesca a ricordare gli eventi del giorno precedente ma che non riesca mai ad intrattenere una connessione diretta con gli eventi dei giorni prima. Nell'ottica di Parfit anche in questo caso si stabilirebbe di giorno in giorno una relazione R che garantirebbe il permanere dell'identità di questa persona, nonostante questa non sia in grado di ricordare nulla del suo passato precedente al giorno prima esattamente come se non si stabilisse la relazione R. Come può però la relazione R essere "ciò che conta" se in questo caso sembra essere irrilevante? Una possibile replica dal punto di vista di Parfit sarebbe sostenere che "ciò che conta" è la relazione R solo se accompagnata da una connessione diretta. In questo modo però la teoria di Parfit si riavvicina pericolosamente a quella lockeana in cui si sosteneva che ciò che fa davvero la differenza nel nostro sopravvivere non è una qualche forma di continuità psicologica ma la possibilità di una effettiva connessione diretta.

Consideriamo infine il fatto che il mio io attuale non può essere collegato da una relazione R al mio io che sviluppò le mie capacità di categorizzazione, di riconoscimento delle cose, delle persone e più in generale del mio io infante che ha sviluppato tutte quelle abilità tipiche anche della mia psicologia di adulto che considero

rilevanti nel definire la mia personalità. La neuropsicologia inoltre ci insegna che alcuni traumi possono cancellare in parte o completamente la mia memoria a breve o a lungo termine, mantenendo però intatte altre forme di memoria procedurale grazie alla quale chi ha subito il trauma può continuare ad essere in grado di parlare la sua lingua madre o di suonare il violino ad esempio. In casi come questi sembra quindi che nonostante il trauma distrugga tutti i requisiti necessari al sussistere della relazione R, ciò nonostante vengano mantenuti da queste persone alcune capacità che sono una parte essenziale della loro identità personale.

2.4 Conclusioni.

Ho cercato in questo capitolo di presentare alcune tra le argomentazioni fondamentali sulle quali si basano le diverse visioni che hanno caratterizzato il dibattito sull'identità personale dalla seconda metà del novecento ad oggi, in modo da fissare i requisiti teorici per quanto segue. Le tre macro-visioni che sono state considerate in maniera spero abbastanza chiara anche se sintetica e parziale, non rappresentano la totalità delle teorie formulate fino ad oggi. Alcuni filosofi, come ad esempio Agar (2004), hanno cercato di formulare prospettive definite *ibride* per il loro tentativo di conciliare criteri psicologici e fisici, nel caso di Agar inserendoli in un ottica di tipo funzionalista. Avendo però questo capitolo l'unico scopo di dare una visione d'insieme del dibattito per come si è delineato fino ad oggi ho scelto per ragioni di spazio di considerare solo le principali prospettive in cui è stato inquadrato il problema dell'identità personale. Anche la scelta degli autori, soprattutto nel caso della parte sul criterialismo psicologico, è stata fatta in maniera funzionale a quanto seguirà nel prossimo capitolo nel quale finalmente verranno trattati direttamente gli argomenti ed i temi dai quali derivano gli scopi di questo testo.

3.

IL CRITERIALISMO PSICOLOGICO ED IL CONCETTO DI IDENTITÀ'.

Tra le obiezioni rivolte sia alla teoria di Shoemaker, sia a quella di Parfit che sono state considerate nell'ultima parte del secondo capitolo, una categoria particolare è rappresentata da quelle conseguenti ai casi di fissione o a particolari tipi di teletrasporto, ovvero i casi in cui l'identità personale sembra prendere una forma ramificata. Questo ordine di obiezioni si è rivelato particolarmente difficile da superare per chi sostiene che la continuità psicologica sia un criterio sufficiente per l'identità personale, abbiamo ad esempio potuto notare che Parfit nel definire in base alla relazione R l'identità personale è stato costretto ad aggiungere una clausola ad hoc per bloccare le conseguenze teoriche della ramificazione della relazione R.

Il problema fondamentale rappresentato da questo tipo di casi per chi difende la sufficienza del criterio psicologico è il fatto che si verifica la presenza di più candidati ugualmente validi per l'identità personale. Non potendo due soggetti tra loro diversi essere entrambi identici ad uno di partenza, molti criterialisti psicologici per difendere la loro teoria si sono trovati costretti a difendere un concetto di identità che violasse quello che viene chiamato "*the only x and y principle*", ovvero il principio secondo cui per stabilire se sussiste la relazione di identità tra due oggetti, non si deve prendere in considerazione nessun fatto che sia al di fuori dei due oggetti stessi. Nel caso in cui io volessi stabilire se due persone sono identiche, secondo il principio sopra riportato, non dovrò considerare nessun fatto ulteriore all'infuori della loro continuità fisica, psicologica o qualsiasi altro fatto unicamente relativo a quelle due persone. Per far fronte ai casi citati sopra mantenendo ferma la sufficienza del criterio psicologico di identità però, alcuni filosofi come Parfit (1984) e Nozick (1981) hanno difeso concezioni di identità che, escludendo il *only x and y principle*, ammettono la plausibilità dell'esistenza di candidati migliori o peggiori per l'identità personale i quali possono aver bisogno dell'attuarsi di condizioni relative al contesto esterno perché venga stabilita la relazione di identità o meno, come ad esempio il fatto che il teletrasporto dopo aver creato la copia distrugga l'originale.

Ovviamente questa categoria di formulazioni non è stata risparmiata dalle critiche, in particolare in questo testo citerò quelle formulate da Wiggins (2001) e Noonan (2003) che, seppure da prospettive radicalmente differenti, hanno entrambi difeso una visione dell'identità assoluta e determinata.

3.1. La dipendenza sortale dell'individuazione e la conseguente assolutezza e determinatezza del concetto di identità.

La tesi della dipendenza sortale dell'individuazione è stata difesa da David Wiggins (2001) in contrapposizione ad un'altra tesi che ne sosteneva invece la relatività. Attraverso la formula "dipendenza sortale" Wiggins si riferisce principalmente a due condizioni: la prima è che qualsiasi cosa esista deve necessariamente appartenere ad una categoria individuativa o basilare, ovvero è un qualche cosa, e capire pienamente che cosa significhino frasi come "quell'oggetto corre" implica capire che cosa sia quell'oggetto. La seconda condizione è che nel momento in cui qualcuno afferma che la cosa che corre è la stessa cosa che è bianca, ad esempio, questo giudizio non può essere vero a meno che la cosa che è bianca non appartenga alla stessa categoria individuativa della cosa che corre e che la cosa in questione sia associata ad un principio attraverso il quale entità di un certo tipo vengono tracciate e reidentificate con successo come le stesse entità. La tesi della relatività dell'identità invece rappresenta l'ipotesi secondo la quale a domande come "Che cosa era quella cosa?" si possono dare risposte sortali diverse e ugualmente corrette. Secondo questa teoria ai sensi della reidentificazione di un oggetto attraverso il tempo è rilevante distinguere sotto quale concetto stiamo sussumendo quell'oggetto. Nel caso in cui l'oggetto in questione sia un essere umano ad esempio, ai fini della sua reidentificazione sarebbe in quest'ottica rilevante il fatto che noi stiamo considerando l'essere umano come uno studente in un dato tempo e come un soldato in uno successivo, in modo tale che uno stesso particolare possa coincidere con un altro se sussunto sotto un determinato concetto e non coincidere se invece considerato sotto un altro.

La tesi che viene sostenuta da Wiggins nei confronti della relatività dell'identità è che nonostante uno stesso particolare possa essere individuato attraverso diversi e ugualmente validi concetti sortali, questi rappresentano soltanto molteplici modi per arrivare alla stessa conclusione e dare la stessa risposta alla domanda *lo stesso cosa?*. Quello che Wiggins sta sostenendo è che la riflessività e la congruenza dell'identità implicano che: se a è lo stesso qualcosa di b , lo stesso albero, cavallo, eccetera, allora tutte le diverse procedure di individuazione di a , se realmente individuano a , portano necessariamente alla stessa risposta relativamente alla coincidenza di a con b . Per dimostrare la sua tesi Wiggins formalizza in questo modo il suo requisito:

$$((\exists f) \left(\begin{smallmatrix} a = b \\ f \end{smallmatrix} \right)) \rightarrow ((\forall g)(g(a) \rightarrow \left(\begin{smallmatrix} a = b \\ g \end{smallmatrix} \right))).$$

Per dimostrare la falsità della tesi della relatività dell'identità Wiggins utilizza come primo assioma l'estensione della forte riflessività dell'identità per cui $(\forall x)(x = x)$, al concetto di *essere la stessa f*, per cui si avrà:

$$(\forall x)(f(x) \rightarrow \left(\begin{smallmatrix} x = x \\ f \end{smallmatrix} \right)).$$

Si serve inoltre della Legge di Leibniz, ovvero l'Indiscernibilità degli identici secondo cui: $(\forall x)(\forall y)((x = y) \rightarrow (\phi x \leftrightarrow \phi y))$ dal quale similmente a quanto fatto precedentemente deriva:

$$(\forall x)(\forall y)\left(\begin{pmatrix} x = y \\ f \end{pmatrix} \rightarrow (\phi x \leftrightarrow \phi y)\right).$$

Trattando la lettera f non come una lettera schematica ma come una variabile Wiggins ottiene quindi:

$$\begin{aligned} &(\forall x)(\forall y)((\exists f)\left(\begin{pmatrix} x = y \\ f \end{pmatrix} \rightarrow (\phi x \leftrightarrow \phi y)\right)) \\ &(\forall x)(\forall y)(\forall f)\left(\begin{pmatrix} x = y \\ f \end{pmatrix} \rightarrow (\phi x \leftrightarrow \phi y)\right). \end{aligned}$$

A questo punto si consideri il fatto che la tesi della relatività dell'identità implica che dati a , b , f e g si abbia:

$$\begin{pmatrix} a = b \\ f \end{pmatrix} \text{ e non } \begin{pmatrix} a = b \\ g \end{pmatrix} \text{ e } g(a).$$

Se però si prende la prima parte $\begin{pmatrix} a = b \\ f \end{pmatrix}$ come vera e si considera la ϕ nello schema della legge di Leibniz $\begin{pmatrix} a = b \\ f \end{pmatrix} \rightarrow (\phi a \leftrightarrow \phi b)$ come comprendente tra i suoi sostituti qualsiasi predicabile della forma $\begin{pmatrix} a = x \\ g \end{pmatrix}$ a questo punto avremo:

$$\begin{pmatrix} a = b \\ f \end{pmatrix} \rightarrow \left(\begin{pmatrix} a = a \\ g \end{pmatrix} \leftrightarrow \begin{pmatrix} a = b \\ g \end{pmatrix}\right).$$

Ma per *modus ponens* e la assunzione iniziale $\begin{pmatrix} a = b \\ f \end{pmatrix}$ conseguentemente avremo:

$$\begin{pmatrix} a = a \\ g \end{pmatrix} \leftrightarrow \begin{pmatrix} a = b \\ g \end{pmatrix}.$$

A questo punto per la riflessività dell'identità e per il terzo congiunto della tesi della relatività dell'identità $g(a)$ abbiamo $\begin{pmatrix} a = a \\ g \end{pmatrix}$ e quindi nuovamente per *modus ponens*:

$$\begin{pmatrix} a = b \\ g \end{pmatrix}.$$

Wiggins in questo modo è arrivato alla negazione della tesi della relatività dell'identità basandosi unicamente sulla riflessività della relazione di identità e sulla Legge di Leibniz che Wiggins difende dai suoi detrattori sostenendo che i controesempi che sono stati riportati rispetto ad essa sono relativi a contesti modali o intensionali e quindi scarsamente interessanti. Inoltre nel caso in cui questa definizione venga abbandonata, bisognerebbe trovarne un'altra che preservi universalmente la sostituibilità degli identici in quanto principio evidentemente valido.

Per avvalorare ulteriormente la tesi della non relatività del concetto di identità Wiggins successivamente alla sua dimostrazione, considera cinque casi in cui la relatività dell'identità sembra essere esemplificata. I primi due casi, liquidati rapidamente come irrilevanti, sono situazioni in cui g è semplicemente un concetto sotto il quale non possono essere sussunti né a né b , come ad esempio nel caso di: la stella del mattino è *la stessa stella* della stella della sera; in quanto Venere non è una stella ma un pianeta. Oppure casi in cui semplicemente non esiste g per cui a è lo stesso g di b , come nel caso di Venere è la stessa stella di Marte.

I successivi tre casi che Wiggins considera meritano invece più profonda attenzione. Il primo è quello in cui John Doe, il ragazzo che era stato preso per stupido, è *lo stesso essere umano* di Sir John Doe, sindaco di Londra, ma non è *lo stesso ragazzo*, in quanto Sir John Doe non è un ragazzo, né *lo stesso sindaco* o *lo stesso padre di cinque figlie*, in quanto il ragazzo non aveva un lavoro né aveva avuto figli. In questo caso però ragazzo, sindaco, padre, eccetera sono tutte parole sortali che si riferiscono a concetti validi in questa situazione. Apparentemente quindi questo sembra essere un caso in cui:

$$\left(\begin{smallmatrix} a = b \\ f \end{smallmatrix} \right) \text{ e non } \left(\begin{smallmatrix} a = b \\ g \end{smallmatrix} \right) \text{ e } (ga \vee gb) \text{ e non } (gb).$$

Ovvero un caso in cui si verifica l'identità tra a e b sotto un concetto sortale f (essere umano), mentre invece non si realizza sotto un altro g (ragazzo). Questa categoria di casi ad ogni modo non è esattamente quella di cui sono alla ricerca i relativisti. L'apparente relatività dell'identità di queste situazioni deriva principalmente dalla confusione tra affermazioni atemporali e proferimenti in cui invece il tempo dei verbi deve essere preso sul serio ai fini della loro verifica. In questo caso l'affermazione "Sir John Doe è lo stesso ragazzo di John Doe" viene plausibilmente rifiutata in quanto Sir John Doe non è più un ragazzo, ma al contrario l'affermazione "Sir John Doe è *stato* lo stesso ragazzo di John Doe" risulta essere assolutamente non problematica. Il problema risiede nel fatto che "ragazzo" è un termine sortale temporalmente delimitato.

Questa categoria di casi ha portato alla luce la necessità di distinguere tra concetti sortali che si applicano al presente indistintamente ad un individuo durante tutto l'arco della sua esistenza e concetti, come ragazzo o sindaco, che invece non lo sono. Wiggins chiama questa seconda categoria "phased-sortals" e afferma che nonostante possano dare la falsa impressione che a possa essere la stessa f di b ma non la stessa g , non sono in conflitto con la tesi della dipendenza sortale del concetto di identità. Tesi che Wiggins riformula in questo modo: per ogni x e ogni y , ogni concetto che individua adeguatamente x per ogni estensione della sua esistenza comporta la stessa risposta, direttamente o attraverso il principio di individuazione, per il sostantivo a cui si

riferisce, di qualsiasi altro concetto che individua appropriatamente x o y alla domanda se x coincida con y o meno.

La quarta categoria di casi in cui sembra verificarsi la tesi della relatività dell'identità è quella che può essere formalizzata in:

$$\begin{pmatrix} a = b \\ f \end{pmatrix} \text{ e non } \begin{pmatrix} a = b \\ g \end{pmatrix} \text{ e } (g(a) \vee g(b)) \text{ e } (g(a) \text{ e non } g(b)).$$

Questa situazione può essere esemplificata attraverso frasi come: “I cocci di ceramica che vedi laggiù sono il vaso che hai visto ieri”. I cocci in questo caso non possono essere lo stesso vaso, ma possono essere lo stesso insieme di frammenti di materia. Se inoltre riutilizzassimo i cocci che componevano il vaso per fare una caffettiera ad esempio avremmo prodotto una situazione in cui si potrebbe dire che “il vaso è la caffettiera”. Questo è vero se i due oggetti vengono considerati come *la stessa collezione di materia*, mentre è falso se vengono considerati come *lo stesso utensile*. Secondo Wiggins in casi come questo l'origine della confusione risiede nel significato del termine “sono” nella prima frase. Mentre i relativisti considerano il significato dell'è in: “è la stessa collezione di frammenti di materia” come quello di un'uguaglianza infatti, secondo Wiggins il suo significato è comparabile a quello che troviamo in frasi come: “quella porta è legno di ciliegio”. In quest'ottica quindi il concetto “collezione di frammenti di materia” introdotto dalla copula ha la funzione di presentare ciò che *costituisce* l'oggetto a cui si fa riferimento e non quella di essere un concetto sortale sotto il quale sussumere il soggetto della frase nei termini di una sua identità con un altro oggetto. Se infatti supponiamo che il vaso sia lo stesso insieme di cocci di ceramica della collezione di frammenti di materia, allora siamo autorizzati a inferire che il vaso sia un insieme di cocci di ceramica, ma a questo punto secondo il principio per cui se x è ϕ , allora esiste una qualche cosa- ϕ con cui x è identico, dovrà esistere un qualche insieme di cocci di ceramica con cui il vaso è identico. Se però due oggetti sono identici questi devono condividere la stessa storia di esistenza, ma se nel caso in cui io distrugga il vaso distruggessi anche l'insieme di cocci di ceramica allora la frase di partenza “i cocci di ceramica che vedi laggiù sono il vaso che hai visto ieri” sarebbe falsa e non potrebbe qualificarsi come un'esemplificazione della relatività dell'identità. Se invece nel distruggere il vaso io non distruggessi anche l'insieme di cocci di ceramica allora non potrebbe essere vero che il vaso è un insieme di cocci di ceramica, perlomeno non nel senso di *è identico a*, ma si starebbe invece affermando che il vaso è costituito da ceramica.

La quinta categoria di casi in cui sembra verificarsi la tesi della relatività dell'identità viene da Wiggins così formalizzata:

$$\begin{pmatrix} a = b \\ f \end{pmatrix} \text{ e non } \begin{pmatrix} a = b \\ g \end{pmatrix} \text{ e } (g(a) \vee g(b)) \text{ e } (g(a) \text{ e } g(b)).$$

Questa formula fa riferimento ad esempio al fatto che se io facessi il bagno in un fiume in due giorni differenti io farei il bagno nello stesso fiume ma non nella stessa acqua, nonostante i fiumi siano acqua infatti, il fiume in due tempi diversi non è la stessa acqua. Similarmente il ragazzo John Doe è lo stesso essere umano di Sir John

Doe ma non è lo stesso aggregato di cellule nonostante un essere umano sia un aggregato di cellule. Un caso diverso è quello è quello rappresentato da enunciati come: “Il sindaco non è lo stesso funzionario dell’amministratore delegato della Fiat, infatti si scrivono l’un l’altro lettere, ma è lo stesso uomo.” Questo esempio è tratto dal libro di Peter Geach (1962, pag. 157) *Reference and Generality*.

Per quanto riguarda il caso del fiume notiamo facilmente che l’interpretazione dell’enunciato che identifica lo stesso fiume attraverso il predicato sortale è *la stessa acqua* non è compatibile con la legge di Leibniz in quanto non tutte le proprietà dell’uno appartengono anche all’altra e viceversa, il fiume in cui ho ipoteticamente fatto il bagno ieri ad esempio non è ora parte del mare mentre l’acqua sì. La “stessa acqua” semplicemente non è un concetto sortale in grado di sussumere un enunciato di identità che identifichi un fiume con qualcosa. Nel caso dell’identificazione di John Doe attraverso il concetto “stesso aggregato di cellule” notiamo che, secondo la legge di Leibniz, se John Doe viene considerato identico ad un aggregato di cellule allora tutte le proprietà dell’uno devono essere anche proprietà dell’altro. Mentre però, a causa del continuo ricambio cellulare, ad un aggregato di cellule se ne sostituisce un altro, alla persona John Doe non se ne sostituisce similmente un’altra. Come nel caso del vaso frantumato citato precedentemente, anche in questo esempio secondo Wiggins “aggregato di cellule” deve in realtà essere interpretato in maniera costitutiva, se infatti interpretassimo “aggregato di cellule” come “composto formato da una successione di costituenti continuamente rinnovata attraverso il tempo”, allora il ragazzo John Doe e l’uomo John Doe sarebbero effettivamente lo stesso aggregato di cellule e questo non apparirebbe più come un esempio della relatività dell’identità.

Nel caso dell’esempio tratto dal libro di Geach (1962) invece, da un certo punto di vista è semplicemente falso che il sindaco e l’amministratore delegato non siano lo stesso funzionario. L’interpretazione che rende la prima parte dell’enunciato vera è quella che concerne quali siano le funzioni dei due uffici e cosa significhi assolvere quelle funzioni. Ciò che realmente sta affermando l’enunciato “il sindaco non è lo stesso funzionario dell’amministratore delegato della Fiat, ma è lo stesso uomo” è quindi che sindaco e amministratore delegato della Fiat sono due concetti diversi e che soddisfare uno dei due non comporta necessariamente soddisfare l’altro. L’aggiunta della parte “infatti (loro) si scrivono l’un l’altro delle lettere” comporta la formazione di uno zeugma, la parola “loro”, sottintesa nella frase, per come viene utilizzata infatti può essere interpretata come: “i due funzionari”, o come: “l’uomo che soddisfa il primo concetto e l’uomo che soddisfa il secondo concetto” ed è in questa seconda interpretazione che “loro” contingentemente sono lo stesso funzionario e lo stesso uomo.

Wiggins in tutte le tipologie di casi prese in considerazione ha fatto risalire l’apparente relatività sortale dell’identità ad una confusione dovuta all’ambiguità del significato di alcuni usi della copula “è”, oppure all’ambiguità di alcune espressioni referenziali nel designare il loro referente, come nei casi in cui uno stesso nome può riferirsi sia all’oggetto sia alla materia che lo costituisce. Dopo aver risolto i casi che apparivano più problematici per la sua tesi della dipendenza sortale dell’identità Wiggins è in grado di affermare congiuntamente e di formalizzare l’assolutezza dell’identità e il suo principio della dipendenza sortale, in particolare partendo dalla Legge di Leibniz:

$$x = y \rightarrow (\varphi x \leftrightarrow \varphi y)$$

e dalla formalizzazione del principio della dipendenza sortale:

$$x = y \text{ se e solo se } (\exists f) \begin{pmatrix} x = y \\ f \end{pmatrix}$$

ottiene lo schema:

$$(\exists f) \begin{pmatrix} x = y \\ f \end{pmatrix} \rightarrow (\varphi x \leftrightarrow \varphi y)$$

Stando a questa formulazione quindi dato un qualsiasi predicato sortale g per cui $\frac{x = y}{g}$, x sarà φ solo se anche y è φ , ovvero:

$$(\forall g) \left(\frac{x = y}{g} \rightarrow \varphi x \leftrightarrow \varphi y \right)$$

Consideriamo ora i due individui x e y entrambi sussunti sotto il concetto sortale g in quanto sono lo stesso g . Se prendiamo un altro concetto sortale h abbiamo che o x appartiene ad h oppure no. Se x non appartiene ad h allora secondo il principio per cui $V \rightarrow (F \rightarrow p)$ abbiamo:

$$\frac{x = y}{g} \rightarrow (h(x) \rightarrow \frac{x = y}{h}).$$

Se invece x appartiene ad h allora si avrà:

$$\frac{x = y}{g} \text{ e } h(x).$$

In questo caso si ottiene che $\frac{x = x}{h}$ e x quindi possiede il predicato $\begin{pmatrix} x = \dots \\ h \end{pmatrix}$.

Secondo la legge di Leibniz e l'eguaglianza $\frac{x = y}{g}$ stabilita precedentemente di conseguenza anche l'individuo y possiede questo predicato e quindi ugualmente a quanto avveniva per l'altra possibilità si ottiene che:

$$\frac{x = y}{g} \rightarrow (h(x) \rightarrow \frac{x = y}{h}).$$

Riformulando questa conclusione utilizzando anche i quantificatori si arriva quindi alla conclusione:

$$(\forall h)(\forall g)(\frac{x = y}{g} \rightarrow (h(x) \rightarrow \frac{x = y}{h})).$$

La quale implica la falsità della tesi della relatività dell'identità.

Attraverso queste argomentazioni Wiggins arriva alla conclusione che necessariamente, se si vuole formulare una teoria dell'individuazione, questa deve essere fondata sull'assolutezza del concetto di identità stabilita dalla legge di Leibniz e sulla sua tesi della dipendenza sortale dell'identità. Mentre a Wiggins queste considerazioni servono da base preliminare per formulare la sua teoria dell'individuazione, in questo contesto molte delle argomentazioni di questo e del prossimo capitolo serviranno da fondamenta teoriche per le critiche che verranno indirizzate contro gli esponenti delle "Closest Continuer Theories".

3.2. L'identità degli artefatti e l'only x and y principle.

Delle Cose che esistono, le une sono da natura, le altre da altre cause. Da natura sono gli animali e le loro parti, le piante e i corpi semplici, come ad esempio la terra, il fuoco, l'aria e l'acqua; infatti e gli altri corpi dello stesso tipo, noi diciamo che sono da natura. E tutte le cose sopra richiamate è chiaro che sono differenti rispetto alle cose che non esistono da natura. E' manifesto, infatti, che tutte le cose che sono da natura, hanno il principio del movimento e del riposo in sé stesse, le une secondo lo spazio, le altre secondo crescita e diminuzione, altre ancora secondo l'alterazione. Invece un letto o un mantello, e ogni altro oggetto di questo genere, in quanto a ciascuno di essi compete questa denominazione – e cioè in quanto essi sono prodotti da tecnica -, non possiede in sé stesso nessuna tendenza innata al cambiamento; ma essi hanno un tale impulso e tanto esteso, solo in quanto sono di pietra o di legno, o di qualcosa di misto; allora la natura è principio e causa dell'essere in movimento e dello stare in riposo di ciò cui essa appartiene originariamente, per sé stessa e non in modo accidentale.

E dico "non in modo accidentale", in quanto potrebbe capitare che costui, essendo medico, sia causa della salute a sé stesso; ma non è in quanto oggetto di guarigione che egli possiede l'arte del medico, ma è capitato accidentalmente a colui che è medico di essere oggetto di guarigione – ed è proprio per questo, che questi due aspetti – di essere cioè medico e paziente – sono separati l'uno dall'altro. Lo stesso accade a tutte le altre cose che sono prodotte artificialmente. Nessuna di esse, infatti, ha in sé stessa il principio della propria produzione. Ma per le une, questo principio è in altro ed esterno, come ad esempio quando si tratta di una casa o di quanto è oggetto di produzione manuale, mentre, per le altre, questo principio è immanente alle cose stesse, ma non appartiene ad esse per sé stesse, come ad esempio nel caso di quelle cose che potrebbero causare, in modo accidentale, un mutamento in sé stesse. (Aristotele 192b 8-34).

Quella che Aristotele nel primo capitolo del secondo libro della *Fisica* sta tracciando è una distinzione fondamentale tra oggetti naturali e artefatti, attribuendo agli

uni un principio di attività o movimento, che gli altri possono possedere solo accidentalmente e secondariamente a causa del loro essere composti da una sostanza naturale. Questa distinzione fondamentale ha delle ripercussioni che concernono i criteri ed i requisiti che una buona teoria dell'identificazione deve riuscire a soddisfare. Mentre in generale sembra possibile individuare gli oggetti naturali facendo riferimento al loro principio di attività innato, gli artefatti vengono individuati in quanto sono una determinata quantità di materia organizzata in modo tale da svolgere una funzione. Questa caratteristica dell'individuazione degli artefatti comporta notevoli differenze rispetto all'individuazione degli oggetti naturali. È ad esempio perfettamente comprensibile che un orologio si fermi per un dato periodo di tempo, in quanto scarico o rotto, per poi riprendere a funzionare continuando a persistere come lo stesso orologio. Il concetto di "orologio" deve quindi comprendere in qualche modo la funzione dell'indicare l'ora, senza però richiedere, a differenza di quanto avviene per gli oggetti naturali, che questa funzione venga svolta in maniera continua. L'unico modo in cui può venire compromessa la persistenza di un orologio è la perdita irreversibile della capacità di svolgere la funzione di indicare l'ora. Un'altra ragione per cui le condizioni di persistenza di un artefatto non devono essere troppo strette è il fatto che uno stesso artefatto può venire smantellato e riassembleato nuovamente, se inoltre alcune delle sue parti sono danneggiate, ad esempio, queste possono venire sostituite senza che per questo venga meno l'identità dell'oggetto intero.

Stando così le cose però sembra che una volta accomodati tutti i requisiti di funzionamento discontinuo, di sostituzione delle parti e di smantellamento e riassettaggio, la teoria dell'individuazione che ne rimane sia così povera da non poter prevenire che un artefatto normalmente identificato ad un tempo t' possa venire identificato come coincidente al tempo t'' con due artefatti distinti conseguenti ad esempio ad una riparazione radicale ed al riutilizzo di alcune sue parti, conseguenza che qualsiasi teoria dell'identificazione dovrebbe voler scongiurare.

L'esempio più classico che viene usato per trattare questo tipo di casi è quello che venne utilizzato a questo scopo da Hobbes e tratto dalla "Vita di Teseo" di Plutarco che riguarda la nave di Teseo. In questo esempio si suppone che la nave di Teseo, nel corso degli anni subisca continue sostituzioni di parti fino ad essere interamente composta di parti non originali. Si suppone inoltre che tutte le parti che sono state sostituite siano state conservate in un magazzino e che siano state poi rimontate insieme esattamente nello stesso ordine per formare nuovamente la nave. Stando a questo esempio sia la nave composta unicamente di parti sostituite sia quella riassembleata sembrano essere entrambe numericamente identiche con l'originale nave di Teseo, cosa assurda.

Lo scopo originario per il quale Hobbes ha formulato questo esempio era quello di criticare chi definiva il *principium individuationis* come "unità di forma". Questa visione viene considerata da Wiggins come non molto dissimile dalla sua ipotesi della dipendenza sortale dell'identità, senonché stando alla sua formulazione, sembrerebbe da preferirsi la nave riparata perché persistente e disponibile a Teseo senza interruzioni per tutto l'arco di tempo considerato. Sotto altri aspetti però sembra che a dover essere considerata come la stessa nave sia la nave ricostruita, questa infatti è composta della stessa materia organizzata nello stesso ordine e probabilmente è più simile all'originale di quella riparata sotto il punto di vista dell'aspetto esteriore o della manovrabilità ad esempio. Una possibile soluzione del problema consiste nel supporre che la confusione

risieda nel significato della parola “nave”; un capitano ed un antiquario infatti pur riferendosi alla nave di Teseo utilizzando lo stesso termine “nave”, nell’utilizzarlo hanno interessi diversi e sono quindi portati a fare una scelta diversa riguardo a quale sia la “stessa nave” rispetto all’originale nave di Teseo. L’antiquario ed il capitano avendo interessi diversi starebbero quindi utilizzando la parola “nave” con due significati diversi e quindi anche con due principi di identificazione diversi. Questa soluzione concettualista del problema però è in realtà altrettanto poco soddisfacente sia quanto quella mereologica, che si suppone abbia utilizzato l’antiquario nell’identificare la nave riasssemblata come la stessa nave della nave di Teseo, sia quanto quella utilizzata dal capitano basata acriticamente sulla continuità della persistenza.

Un quarto tentativo di soluzione al problema è quello che consiste nell’ammettere la possibilità che esistano in questi casi dei candidati più o meno validi per l’identità degli oggetti, elaborando una teoria che, in base a determinati criteri, attribuisca diversi gradi di forza alla candidatura. Nello specifico, in quest’ottica x coincide con y sotto il concetto f se e solo se x è un f -candidato all’identità con y e non esiste un altro f -candidato altrettanto valido all’infuori di x all’identità con y . Se applichiamo questa definizione al caso della nave di Teseo avremo che: nel caso in cui la nave originale venga continuamente riparata e le sue parti sostituite, ma senza che queste vengano riutilizzate, la nave riparata deve essere considerata come la stessa nave. Nel caso in cui invece ad esempio, la nave riparata sia stata affondata e sia rimasta solo quella ricostruita, allora la nave ricostruita è da considerarsi come la stessa nave della nave di Teseo. Se infine entrambe le navi continuano a sussistere come candidati per l’identità con la nave di Teseo, allora sarà da preferirsi quella riparata perché continuamente persistente.

Questa attualizzazione della teoria che d’ora in avanti chiamerò *Teoria del miglior candidato* mostra subito la sua più grande contraddizione: sembra infatti inconcepibile che l’identità numerica di due oggetti x e y possa dipendere da elementi contestuali e contingenti esterni ad x e y stessi. Questo principio che viene chiamato “*only x and y principle*”, non intende negare che la singolarizzazione di un oggetto x e di un oggetto y possa dipendere anche da elementi contestuali a loro esterni, ma stabilisce che per verificare la coincidenza o non coincidenza di un oggetto x con un oggetto y , oppure per ripercorrere la storia passata di x per ricostruire il collegamento di identità con y , gli unici fatti di cui devo preoccuparmi sono quelli riguardanti x e y stessi, ovvero a loro intrinseci.

La soluzione proposta dalla Teoria del miglior candidato al problema dell’identità della nave di Teseo viola apertamente questo principio facendo dipendere la scelta tra i due candidati all’identità da fatti contingenti come ad esempio la mancata ricostruzione di una nave con i pezzi sostituiti o l’affondamento della nave riparata. Anche nel caso in cui sussistano contemporaneamente sia la nave riparata sia quella ricostruita, sostenere che la nave riparata sia un miglior candidato per l’identità con la nave di Teseo rispetto alla nave ricostruita implica infrangere l’*only x and y principle*. La scelta tra le due è infatti in quest’ottica fondata sulla maggiore forza o debolezza delle ragioni dei due candidati considerati contemporaneamente. Quello che sembra paradossale stando alla Teoria del miglior candidato, è il conseguente passaggio modale: “La nave ricostruita *sarebbe stata* identica alla nave di Teseo se la nave riparata *fosse* affondata”. Secondo il concetto di identità numerica che siamo abituati ad utilizzare è assurdo affermare che

due oggetti numericamente differenti *sarebbero stati* lo stesso oggetto se si *fossero* attualizzate circostanze differenti. L'elemento teorico che necessariamente porta all'infrazione dell'*only x and y principle* sembra quindi essere l'ammissione della possibilità stessa dell'esistenza di candidati rivali per l'identità. Ciò che rende plausibile a prima vista questa possibilità però probabilmente è la confusione tra una concezione comune e stretta di identità, con una nozione invece larga che siamo soliti utilizzare nei confronti delle condizioni di persistenza degli artefatti. Per come è stata fino ad ora esposta riguardo all'identità degli artefatti, la Teoria del miglior candidato non sembra quindi in grado di difendere intellegibilmente il suo tentativo di sostenere la plausibilità dell'esistenza di rivali per l'identità di un oggetto.

In realtà noteremo più avanti nel testo che le cose non sono così semplici e che alcune versioni della Teoria del miglior candidato, come ad esempio quella proposta da Nozick, non saranno così facilmente screditabili. Se infatti il *only x and y principle*, viene inserito da Wiggins tra gli assiomi della sua teoria dell'identificazione con lo scopo di escludere la possibilità dell'esistenza di candidati rivali per l'identità di un oggetto, stando semplicemente alla formulazione generale che ne abbiamo dato sopra, non è chiaro come questo possa necessariamente escludere ogni possibile formulazione della teoria del miglior candidato. Il *only x and y principle*, infatti dispone semplicemente che se x al tempo t' è identico a y al tempo t'' questo dipende esclusivamente da fatti che riguardano solo x e y e dalla relazione intrinseca tra di loro. Se supponiamo però che sia vero che in una situazione possibile x sia identico a y mentre in un'altra x sia identico a z , i fatti che vengono presi in considerazione nella seconda situazione possibile semplicemente non sono fatti che riguardano qualcosa d'altro rispetto a x e y e ciò nonostante avremmo scelto il miglior candidato.

Questo però non significa che *only x and y principle* e Teorie del miglior candidato siano in fondo compatibili, per aggirare l'ostacolo infatti, Robert Nozick è costretto a considerare gli individui come *vermi quadrimensionali* composti di parti temporali, su questo e sui problemi che questa ipotesi comporta però ritornerò più esaurientemente nelle parti finali di questo capitolo. È importante specificare inoltre che Wiggins inserisce questo principio dopo aver già stabilito formalmente che una soddisfacente teoria dell'identificazione deve avvalersi di un concetto di identità che renda conto del suo essere assoluta, transitiva, riflessiva e simmetrica come risulta dalla definizione di Leibniz. L'ontologia di riferimento di Wiggins, più tradizionalmente, è invece un'ontologia tridimensionalista che non attribuisce agli oggetti parti temporali, ma solo le tre dimensioni a loro ascritte dal senso comune, gli oggetti in questo senso sono *permanenti* o *continuanti* e sono interamente presenti in ogni momento della loro esistenza. Considerare le entità come "interamente presenti" in ogni momento della loro esistenza significa che le entità mantengono permanentemente tutte le loro proprietà; questo però non significa che se un oggetto ha la proprietà di "essere verde" al tempo t' necessariamente deve conservare la proprietà di "essere verde" al tempo t'' , significa invece che necessariamente conserverà per tutto l'arco della sua esistenza la proprietà di "essere verde al tempo t' ". Conseguentemente Wiggins, nello stabilire se un oggetto precedente x sia identico ad un oggetto successivo y e per tracciare retrospettivamente la storia delle connessioni tra i due individui, considera come fatti riguardanti x e y la totalità delle loro proprietà e non una limitazione di questo insieme ristretto a quelle di una loro parte temporale riferita solo ad una delle situazioni possibili. Nell'ottica di una

ontologia tridimensionalista tradizionale quindi l'ammissione dell'esistenza di candidati per l'identità comporta effettivamente l'infrazione dell'*only x and y principle*, con tutte le conseguenze assurde che implica l'introduzione di fatti esterni e contingenti.

Molte delle difficoltà che caratterizzano la descrizione dell'identità degli artefatti sono comuni a quelle che presenta l'identità personale in alcuni casi particolari, come quello della separazione dei due emisferi del cervello o quello del teletrasporto in stile *Star Trek* citati nell'ultima parte del secondo capitolo. Alcune conseguenze teoriche del Criterialismo Psicologico applicato a quei casi particolari hanno infatti portato alcuni filosofi, come Sidney Shoemaker e Derek Parfit, a difendere una versione della Teoria del miglior candidato e conseguentemente, per evitare i problemi causati dall'infrazione dell'*only x and y principle*, ad adottare un'ontologia quadrimensionalista che suddivide le persone in parti temporali o *Person-Stages* come sono state chiamati. Le conseguenze di una ontologia di questo tipo però, se applicata all'ambito dell'identità personale, si estendono ad ambiti estranei a quelli dell'identità degli artefatti. Come è stato sottolineato fin dalle prime trattazioni Lockeane infatti, tra gli scopi principali della formulazione della sua teoria dell'identità personale c'era la volontà di rendere conto del fatto che la nostra sopravvivenza, così come il nostro passato e futuro, non possono esserci indifferenti. Questa assunzione di base però è esattamente ciò che arriverà a negare Derek Parfit spingendo alle estreme conseguenze la sua Teoria del miglior candidato per l'identità personale, fino a sostenere che l'identità personale non è ciò che conta per la sopravvivenza.

3.3. Le Teorie del miglior candidato e l'identità personale.

Nel suo libro intitolato *Reasons and Persons* (1984) Derek Parfit sostiene una versione del Criterialismo Psicologico che rigetta apertamente l'*only x and y principle* e che prevede l'esistenza di candidati per l'identità personale. La tesi fortemente controintuitiva che viene sostenuta da Parfit quando afferma che l'identità personale non è ciò che conta per la sopravvivenza è che nel momento in cui una persona si preoccupa del suo futuro benessere la sua preoccupazione è derivativa e finalizzata allo scopo del benessere di una persona futura collegata alla presente da una connessione psicologica. La persona futura, che d'ora in poi chiamerò sopravvivente Parfitiano, non necessariamente sarà un sopravvivente letterale e, in un futuro ipotetico in cui saranno possibili trapianti di cervello ed esisteranno macchine per il teletrasporto alla *Star Trek*, una persona potrà assicurarsi di avere un sopravvivente Parfitiano senza dover avere un sopravvivente letterale. Il fatto che in questo momento storico l'unico modo in cui una persona possa assicurarsi di avere un sopravvivente Parfitiano sia quello di avere un sopravvivente letterale è per Parfit puramente contingente e dipendente dalle odierne possibilità tecnologiche. Parfit illustra la sua tesi dicendo che analogamente al fatto che noi diamo un valore ai nostri occhi non per loro stessi ma per la funzione che svolgono, di modo tale che se sapessimo che potrebbero essere sostituiti da un qualche strumento tecnologico che assolvesse pienamente a quella che era la loro funzione, non proveremmo orrore al pensiero della sostituzione dei nostri occhi con un simile

apparecchio, allo stesso modo la nostra preoccupazione di assicurarci un sopravvivate letterale futuro è solo derivativa e quando la tecnologia ci permetterà di assicurarci un sopravvivate Parfitiano, cesserà la nostra preoccupazione. Il ragionamento di Parfit è fortemente controintuitivo perché, ipotizzando di venire a contatto con una popolazione che utilizza come mezzo di trasporto un sistema di teletrasporto in stile *Star Trek*, ovvero un teletrasporto che distrugge tutte le molecole di una persona in un luogo per ricrearle e riordinarle in maniera esattamente identica in un altro realizzando così un suo continuante psicologico perfetto, ci sembrerebbe intuitivamente una pazzia l'utilizzo di una macchina che la popolazione stessa che la usa non considera come un mezzo di trasporto che preserva l'identità di chi lo usa. Quanto sostenuto da Parfit però è esattamente l'opposto, ovvero che sarebbe irrazionale rifiutarsi di utilizzare questo tipo di teletrasporto, a meno di avere qualche interesse particolare nella nostra sopravvivenza letterale, cosa che Parfit nega esplicitamente.

La riflessione di Parfit prende le mosse dalla considerazione di un caso di fissione come quello della separazione e del trapianto separato dei due emisferi del cervello di una persona. Quanto viene sostenuto da Parfit in relazione a casi come questi è che a meno che non sopravviva solo uno dei due prodotti della fissione, la persona originale cessa di esistere letteralmente. Allo stesso tempo però sarebbe alquanto irrazionale, se noi fossimo la persona che sta per subire il trapianto, sia preoccuparci della nostra imminente fissione come se si trattasse della nostra morte, sia pensare di assicurare la nostra sopravvivenza pagando un'infermiera perché distrugga uno dei miei emisferi riducendo il numero dei prodotti della fissione ad uno solo. L'apparente contraddizione di queste due intuizioni del senso comune può per Parfit essere superata solo accettando il fatto che ciò che realmente conta per noi è la nostra sopravvivenza Parfitiana e non quella letterale.

La cornice teorica all'interno della quale Parfit interpreta i casi di fissione è quella di una Teoria del miglior candidato per l'identità personale fondata sulla continuità psicologica. Sostenendo che al momento della fissione la persona originale cessa di esistere e che vengano prodotte due nuove persone infatti viola apertamente l'*only x and y principle* e afferma che l'identità o non identità di una persona con una temporalmente successiva è qualcosa che può essere determinata anche da elementi estrinseci ai due individui. La definizione di sopravvivate Parfitiano che da Parfit è quella rappresentata dall'esistenza dalla connessione psicologica di una persona con un'altra in un tempo successivo, connessione psicologica ottenuta attraverso lo stabilirsi della relazione R, della quale abbiamo già parlato, e del suo stabilirsi nel senso più ampio, ovvero indipendentemente da quale sia la causa per la quale la relazione si stabilisce, quello che conta è che si stabilisca. In questo senso io ho un sopravvivate Parfitiano se esisterà in futuro una persona psicologicamente continua e connessa con me attraverso la relazione R e gli stati psicologici di questa persona derivano in qualche modo dai miei stati psicologici attuali. Il mio sopravvivate Parfitiano sarà inoltre identico a me se sarà l'unico o il miglior candidato all'identità con me in quel dato tempo futuro.

Per illustrare le conseguenze che ha una simile prospettiva sul nostro stesso modo di concepire la nostra sopravvivenza, Noonan (2003) ipotizza un caso in cui dopo la sua morte, una terza persona cada nelle mani di uno scienziato pazzo che ha studiato a menadito la storia di Noonan con l'intento di resuscitarlo. Quello che lo scienziato intende fare è trasformare la terza persona in una replica psicologica di Noonan

utilizzando varie tecniche sperimentali. Nel caso in cui l'operazione avesse successo esisterebbe quindi una persona psicologicamente continua con Noonan che parla e pensa come lui e che crede di essere lui. Secondo la definizione di Parfit questa persona si qualificherebbe indubbiamente come un sopravvivate Parfitiano e da ciò ne deriverebbe che tutte le paure e preoccupazioni che erano state provate dall'originale Harold Noonan di fronte alla morte, sarebbero state infondate in quanto ciò che desiderava realmente non era l'esistenza di un suo sopravvivate letterale ma solo quella di un sopravvivate Parfitiano, che lo scienziato pazzo si è preoccupato di fornirgli. Questo esempio rende evidente quanto sia radicale la proposta di Parfit. Prima di Parfit era infatti largamente accettata tra i sostenitori di un Criterialismo Psicologico per l'identità personale, la tesi secondo cui il fatto che una persona successiva possa considerarsi identica con una precedente con la quale è psicologicamente continua dipende dalla natura della relazione causale che connette gli stati mentali della persona precedente con quelli della successiva. Secondo Parfit invece non importa come si realizzi questa connessione, l'importante è che sussista. Questa conclusione è conseguente al fatto che Parfit attribuisce all'identità personale un'importanza meramente derivativa rispetto alla sopravvivenza. In quest'ottica effettivamente, il fatto che la connessione psicologica con il sopravvivate di una persona si stabilisca attraverso cause affidabili che ne garantiscono l'identità letterale perde di importanza. Se immaginiamo però il caso di un malato terminale che viene posto di fronte alla scelta tra far trapiantare il proprio cervello in un altro corpo o di far creare una sua replica psicologica attraverso un macchinario che registra tutti i suoi ricordi e stati mentali per poi trasferirli nel cervello di un'altra persona, sembra molto plausibile ed intuitivo che il malato terminale scelga il trapianto, e che alla domanda "Perché ha scelto il trapianto?" lui risponda sensatamente "Perché in quel caso io sopravvivo letteralmente, mentre nell'altro no". Stando alla proposta di Parfit però, se l'identità ha importanza solo derivativa per la sopravvivenza, dovremmo intuitivamente considerare le due scelte come equivalenti.

Merito della trattazione di Parfit è stato quello di mettere in evidenza l'impossibilità di sostenere una visione dell'identità che rinunci all'*only x and y principle* e che contemporaneamente mantenga la convinzione del senso comune che l'importanza dell'identità ai fini della nostra sopravvivenza sia non derivativa. Quello che Parfit però non riesce ad ottenere con le sue argomentazioni è l'esclusione necessaria della tesi opposta, ovvero semplicemente quella che mantiene l'*only x and y principle* valido e inamovibile e contemporaneamente fa propria la concezione del senso comune dando all'identità importanza non derivativa. Questa seconda alternativa oltre ad essere notevolmente più vicina alle nostre intuizioni di senso comune sull'identità personale, rimane preferibile alla tesi di Parfit in quanto esistono ragioni forti per considerare valido l'*only x and y principle*, non solo per quanto riguarda l'identità personale, ma per quanto riguarda il concetto generale dell'identità.

L'unico modo in cui Parfit tenta di criticare la plausibilità dell'*only x and y principle* è quello di farne dipendere l'intuitiva validità dalla nostra erronea concezione dell'importanza non derivativa dell'identità per la nostra sopravvivenza. Questa argomentazione però aggira semplicemente il problema, la questione che si sta trattando infatti è se la nostra convinzione che l'identità abbia importanza non derivativa per la nostra sopravvivenza sia infondata o meno. Il problema della argomentazione di Parfit è

che continua a non fornire un valido motivo per cui dovremmo rigettare l'*only x and y principle* e conseguentemente anche le nostre intuizioni del senso comune. Parfit inoltre si sofferma esclusivamente sulla critica della plausibilità dell'*only x and y principle* in relazione al problema dell'identità personale, mentre non compie alcun tentativo di ampliare la portata delle sue argomentazioni all'utilizzo dell'*only x and y principle* inteso come principio valido per il concetto di identità in generale.

Mentre fino ad ora si è solo cercato di mostrare che le argomentazioni di Parfit non escludono necessariamente la concezione del senso comune, Robert Nozick formula anche una obiezione diretta esplicitamente contro la teoria di Parfit. Il punto fondamentale dell'obiezione di Nozick è che l'unico modo in cui Parfit può spiegare la secondarietà della nostra preoccupazione riguardo la nostra sopravvivenza letterale, è quello di intenderla come derivata dalla nostra preoccupazione per la vita ed il benessere futuro di un nostro sopravvissuto Parfitiano. Le intuizioni su cui Parfit fonda la sua tesi però sono in contraddizione con molte altre non meno importanti.

Tenendo presente che la versione del Criterialismo Psicologico adottata da Parfit si basa su una Teoria del miglior candidato per l'identità personale, ovvero considera come condizione *sufficiente* per l'identità di una persona ad un tempo precedente con una ad un tempo successivo il suo essere, in relazione ad una continuità e connessione psicologica, il miglior candidato o, usando la terminologia di Nozick, il *Closest Continuer* esistente. Secondo questo tipo di formulazioni l'esistenza di più di un candidato per l'identità personale con la stessa persona, non implica necessariamente che la persona originale abbia cessato di esistere, questo avviene solo nel caso in cui non sia possibile trovare un miglior candidato, ovvero come lo stesso Nozick dice, nel caso di un pareggio. Lo stesso Parfit considera questo genere di casi denominandoli *Branch Line Cases* ed esemplificandoli come segue:

... I am often teletransported. (con una macchina per il teletrasporto che distrugge le molecole di una persona in un luogo per ricrearle in un altro) I am now back in the cubicle, ready for another trip to Mars. But this time when I press the green button, I do not lose consciousness. There is a whirring sound, then silence. I leave the cubicle, and say to the attendant: "It's not working. What did I do wrong?". "It's working", he replies, handing me a printed card. This reads: "The New Scanner records your blueprint without destroying your brain and body. We hope that you will welcome the opportunities that this technical advance offers". The attendant tells me that I am one of the first people to use the New Scanner. He adds that, if I stay for an hour, I can use the Intercom to see and talk to myself on Mars. "Wait a minute", I reply, "if I'm here I can't also be on Mars". Someone politely coughs, a white-coated man who asks to speak to me in private. We go to his office, where he tells me to sit down, and pauses. Then he says: "I'm afraid that we're having problems with the New Scanner. It records your blueprint just as accurately, as you will see when you talk to yourself on Mars. But it seems to be damaging the cardiac systems which it scans. Judging from the results so far, though you will be quite healthy on Mars, here on Earth you must expect cardiac failure within the next few days". The attendant later calls me on the Intercom. On the screen I see myself just as I do in the mirror every morning. But there are two differences. On the screen I am not left-right reversed. And, while I stand here speechless, I can see and hear myself, in the studio on Mars, starting to speak.

In casi come questo la Teoria del miglior candidato sembra plausibile. Per quanto riguarda la storia appena riportata infatti la persona che rimane sulla terra è il miglior candidato per l'identità con la persona di partenza e quindi è quella persona. Nel caso in cui invece la persona sulla terra fosse stata disintegrata il miglior candidato sarebbe stata la sua replica su Marte che quindi sarebbe stata identica con la persona originale. Anche ammettendo la validità di un simile criterialismo psicologico per l'identità personale, casi come questo fanno sorgere non pochi problemi per la tesi di Parfit. Come viene fatto notare da Nozick infatti, secondo questa visione delle cose, noi ci preoccupiamo del nostro *closest continuer*, quando esiste, indipendentemente da quale sia il grado di continuità che abbiamo con lui. La persona che viene materializzata su Marte infatti ha lo stesso grado di continuità con l'originale sia che questo venga distrutto immediatamente o meno, mentre in un caso però quella persona è il miglior candidato, nell'altro è semplicemente un candidato, semplicemente un sopravvivate Parfitiano. Se proviamo ad immedesimarci nelle due situazioni, quella del teletrasporto tradizionale e quella sopra descritta, e proviamo a pensare a quale sarebbe la nostra attitudine verso il futuro, notiamo che secondo Nozick nella seconda probabilmente avremmo qualche preoccupazione per il nostro sopravvivate Parfitiano su Marte, ma sicuramente ne avremo maggiormente per il nostro *Closest Continuer* rimasto sulla Terra e la notizia dell'imminente attacco di cuore ci farebbe provare una forte paura. Nel caso invece del teletrasporto semplice dovrei essere più preoccupato per l'avvenire del sopravvivate Parfitiano su Marte rispetto a quanto lo sarei stato nel caso precedente; in particolare dovrei essere preoccupato per il suo avvenire quanto lo sono per il mio stesso in ogni situazione. Se le cose stanno così però il grado di preoccupazione che ho verso i miei continuanti non è direttamente proporzionale o determinato dal grado di continuità che ho con loro come viene invece implicato dalla teoria di Parfit.

Personalmente credo che l'intera questione sopra esposta sia un falso problema. Sia Parfit che Nozick infatti pur proponendo tesi fortemente controintuitive, fanno continuamente appello ad ipotetiche "intuizioni del senso comune" per avvalorare le loro argomentazioni. Se a prima vista l'esegesi delle due situazioni, quella del teletrasporto semplice e quella del teletrasporto che preserva il mio corpo sulla terra creando però una disfunzione cardiaca, può apparire coerente, a mio parere esiste un'altra intuizione del senso comune che merita di essere presa in considerazione almeno quanto quelle citate dai due autori e che io personalmente credo sia molto più persuasiva. In entrambe le situazioni, indipendentemente dal fatto che il mio corpo originale muoia di attacco cardiaco o venga disintegrato nel momento esatto in cui io premo il bottone verde del teletrasporto, trovo semplicemente insensato e suicida l'utilizzo di una simile macchina. Nonostante su Marte infatti venga riprodotta una mia replica perfetta e con me psicologicamente continua, nel momento in cui io utilizzo questa particolare macchina l'organismo nato dalla fecondazione di un ovulo di mia madre viene, nel migliore dei casi, disintegrato, io di fatto cesso di esistere. Anche la presunta intuizione del senso comune per cui io dovrei preoccuparmi dell'avvenire del mio sopravvivate Parfitiano è, per me, difficilmente comprensibile. Trovo infatti che questa preoccupazione abbia un'esistenza puramente derivata da cause secondarie come

ad esempio il fatto che, nonostante io non esista più, rimarrà nel mondo qualcuno a fare le mie veci ed a non far mancare il mio affetto alle persone che mi sono care. Personalmente però questa preoccupazione cadrebbe sicuramente in secondo piano di fronte alla prospettiva dell'essere disintegrato, di modo che nel caso io cui io sia nella fase finale di una grave malattia, troverei sì desiderabile assicurare la persistenza di un mio sopravvivate Parfitiano, ma non di certo consolante. Se invece, come nella situazione sopra presentata, io disponessi della possibilità di scegliere tra l'entrare o meno nella macchina del teletrasporto, mai e poi mai sceglierei di entrarci e non mi preoccuperei minimamente dell'avvenire del mio sopravvivate Parfitiano ma esclusivamente del mio. La mia personale comprensione di una situazione simile è talmente lontana da quella di Nozick e Parfit che credo si possa utilizzare come obiezione a chi sostiene che la continuità o connessione psicologica sia un criterio sufficiente per l'identità personale. Come viene messo in risalto da Parfit stesso, infatti è bastata una piccola modifica alla macchina del teletrasporto per rendere più visibile la contraddizione insita nel considerare un sopravvivate Parfitiano come identico alla persona originale. Di fatto è stato sufficiente ritardare di qualche giorno il decesso della persona che entra nella macchina per rendere lei e il suo sopravvivate Parfitiano compresenti nell'universo, di modo che sia impossibile parlare di identità numerica in casi simili, ma al più di identità qualitativa. Questo nonostante tutti i requisiti di continuità e connessione psicologica siano stati mantenuti dal processo di scannerizzazione e riproduzione delle molecole dell'organismo originario.

La forma dell'obiezione di Nozick a Parfit però è convincente. Se infatti la applichiamo al caso della fissione ottenuta dalla separazione dei due emisferi del cervello di una persona e dal trapianto di una delle due metà nel cranio vuoto di un'altra infatti mantiene la sua forza argomentativa. Similmente a quanto avviene nel caso del teletrasporto infatti, sembra che se fossimo certi che entrambi i prodotti della fissione vengano preservati, noi dovremmo preoccuparci maggiormente di quello dei due che conserva anche il nostro corpo in quanto *closest continuer*, relegando l'altro al ruolo di semplice sopravvivate Parfitiano. Nel caso invece in cui fossimo certi che dopo l'operazione l'unico emisfero che continuerà a persistere sarà quello trapiantato nell'altro corpo, allora razionalmente saremmo preoccupati dell'avvenire del nostro sopravvivate Parfitiano in quanto ora è lui il nostro *closest continuer*. Anche se viene applicata ad un caso di fissione notiamo quindi che l'obiezione di Nozick colpisce il segno, contrariamente a quanto sostenuto da Parfit infatti risulta evidente che, nell'ottica di una Teoria del miglior candidato per l'identità personale, il grado di preoccupazione che abbiamo per il futuro dei nostri continuanti non dipende dal grado di continuità che abbiamo con loro.

In conclusione nelle ultime pagine abbiamo potuto notare che esistono due tipi di intuizioni del senso comune riguardanti i casi di *ramificazione*, termine di Parfit, dell'identità personale. Il primo è quello su cui Parfit basa la sua tesi che l'identità non sia importante per quanto riguarda la sopravvivenza, in particolare si tratta dell'intuizione secondo cui un comportamento volto ad assicurare la persistenza di un unico continuante sarebbe insensato, come ad esempio il pagare un infermiere perché distrugga uno dei due emisferi dopo il trapianto. La seconda categoria di intuizioni è invece quella, espressa nell'obiezione di Nozick, che mostra come difficilmente possiamo spiegare la natura della nostra preoccupazione per il futuro dei nostri

continuanti letterali se la intendiamo come semplicemente derivativa dalla nostra preoccupazione per il futuro dei nostri sopravviventi Parfitiani. Abbiamo potuto notare che l'argomentazione di Parfit sulla derivazione della nostra preoccupazione per il futuro, non riesce ad escludere la possibilità di mantenere semplicemente valido l'*only x and y principle*, ma che effettivamente, una volta accettata come valida una Teoria del miglior candidato sull'identità personale, o qualsiasi teoria che rifiuti l'*only x and y principle*, le nostre intuizioni riguardo ai casi di fissione diventano inconsistenti con la visione del senso comune che considera l'identità come fondamentale per la sopravvivenza. Nella prospettiva di una simile Teoria del miglior candidato però l'obiezione di Nozick ha messo in luce il fatto che la tesi di Parfit che ciò che realmente conta per noi è semplicemente la sopravvivenza di un nostro sopravvivate Parfitiano contrasta con il secondo insieme di intuizioni. Stando così le cose sembra che l'unico modo per liberarsi delle contraddizioni sia respingere la teoria di Parfit. Il tentativo di Nozick è quello di fornire una concezione comprensiva e coerente delle nostre intuizioni riguardo ai casi di ramificazione dell'identità, rimanendo all'interno della cornice di una Teoria del miglior candidato che rifiuta l'*only x and y principle*, distinguendo tra la nostra preoccupazione per la nostra futura sopravvivenza letterale e la nostra preoccupazione per il nostro futuro benessere, considerando la prima come effettivamente derivativa rispetto alla nostra preoccupazione per la persistenza di un nostro sopravvivate Parfitiano, mentre la seconda no. Pur sorvolando sul fatto che anche la teoria di Nozick è una Teoria del miglior candidato e perciò è inconsistente con l'*only x and y principle*, vedremo che finirà comunque per scontrarsi con problemi molto simili a quelli con i quali si scontra anche la teoria di Parfit.

3.4. Il tentativo di refutazione dell' *only x and y principle* da parte di Nozick.

Molti sostenitori delle Teorie del miglior candidato per l'identità, tra i quali anche Nozick, difendono il loro punto vista attribuendogli maggiore potenza esplicativa rispetto alle tradizionali teorie dell'identità in relazione alle situazioni enigmatiche che abbiamo incontrato fino ad ora. Per avvalorare la sua tesi e refutare l'*only x and y principle* Nozick propone tra gli altri l'esempio del Circolo di Vienna.

Il Circolo di Vienna era un associazione di studiosi di varie discipline e di estrazione neopositivista che erano soliti riunirsi per discutere di filosofia, logica, matematica e quant'altro prima che l'avvento del nazismo costringesse molti suoi membri di origini ebraiche a fuggire dall'Austria e dalla Germania facendo di fatto scogliere il Circolo. Nozick ci chiede di immaginare che ci fossero originariamente venti membri nel Circolo, dei quali tre fuggirono ad Istanbul e continuarono ad incontrarsi durante gli anni della guerra. Nel 1943 a questi tre arriva all'orecchio la voce che tutti gli altri membri del circolo sono sfortunatamente deceduti e decidono di proclamare le loro riunioni di Istanbul come riunioni del Circolo di Vienna. Nel 1945 però, questi tre vengono informati che altri nove membri del Circolo originale erano riusciti a scappare negli Stati Uniti ed hanno continuato a riunirsi per discutere di logica,

filosofia eccetera. Il gruppo di Istanbul apprende così di essersi fallacemente proclamato “Circolo di Vienna” nel 1943 mentre questo titolo spetta di diritto al gruppo degli Stati Uniti di cui loro sono la filiale di Istanbul. Nozick, riguardo a queste vicende, sostiene che stando alle informazioni che aveva il gruppo di Istanbul aveva ragione sia di proclamarsi Circolo di Vienna nel 1943, sia di riconoscere lo sbaglio ed affermare che il gruppo degli Stati Uniti era il vero Circolo di Vienna nel 1945. Secondo Nozick se mettiamo a confronto le due situazioni, quella solo possibile in cui gli unici membri sopravvissuti sono i tre fuggiti ad Istanbul e quella in cui invece anche altri nove membri sono riusciti ad emigrare negli Stati Uniti, abbiamo il risultato che nella prima situazione l’entità “Circolo di Vienna” viaggia da Vienna ad Istanbul, mentre nella seconda da Vienna agli Stati Uniti ed una seconda entità si forma ad Istanbul o durante il viaggio tra Vienna ad Istanbul. Nozick intende con questo esempio refutare l’*only x and y principle* e mostrare che la sua Teoria del miglior candidato può descrivere consistentemente anche questi casi in cui, stando alle vicende sopra riportate, apparentemente l’identità o meno dell’uno o dell’altro gruppo di studiosi con l’originale Circolo di Vienna dipende estrinsecamente anche da fatti come la sopravvivenza o meno dei nove fuggiti negli Stati Uniti. La risposta di Noonan a questa prima argomentazione è tanto semplice quanto potente. Noonan afferma che tutto quello che c’è da dire di un caso simile è in realtà che nel 1943 i tre studiosi fuggiti ad Istanbul avanzano la pretesa di assumere il titolo di “Circolo di Vienna”, scoprendo poi nel 1945 di aver avuto torto nel farlo. Niente che nella prima situazione abbia viaggiato da Vienna ad Istanbul nella seconda viaggia invece da Vienna agli Stati Uniti, l’unica entità, se realmente ne esiste una, che esiste solo nella seconda situazione nel 1945 si trova negli Stati Uniti, non ad Istanbul.

Nozick sostiene inoltre che la sua *Closest Continuer Theory* sia in grado anche di giustificare consistentemente le intuizioni contraddittorie che abbiamo riguardo ai due scenari descritti da Bernard Williams in *The Self And The Future* (1970) che abbiamo già incontrato e che questa giustificazione allo stesso tempo costituisca una obiezione all’argomento che Williams propone alla fine del testo per avvalorare la sua tesi che il criterio necessario e sufficiente per l’identità personale sia la continuità del corpo.

La prima delle due situazioni descritte nel *Reduplication Argument* di Bernard Williams è quella in cui due persone A e B entrano in una macchina che scambia tutti i tratti psicologici ed i ricordi delle due persone di modo che, al termine del processo, il corpo della persona A abbia i tratti psicologici che erano della persona B e viceversa. L’intuizione, corretta secondo Nozick, che abbiamo di una situazione simile, è che le due persone abbiano scambiato i loro corpi, di modo tale che se prima di entrare nella macchina avessimo chiesto alla persona A, ad esempio, di decidere egoisticamente quale corpo avrebbe preferito che venisse torturato dopo il processo, avrebbe indicato il proprio corpo, perché lui sarebbe stato nell’altro. Nel secondo caso descritto da Williams, troviamo una persona alla quale viene detto che il giorno successivo verrà torturata crudelmente, ma che prima che ciò avvenga le verranno cambiati così radicalmente tutti i tratti caratteriali e psicologici, che potrebbe essere definita un’altra persona. Le intuizioni che abbiamo rispetto a questo caso però contrastano con quelle che avevamo nel primo. Sembra infatti che, non solo il cambio psicologico non possa consolare il soggetto dell’esperimento dalla paura di essere torturato il giorno seguente, ma che anzi il cambio di tratti psicologici possa spaventarli ancora maggiormente

mettendolo anche di fronte alla possibilità di perdere la sua personalità. Se paragoniamo attentamente i due casi però, ci accorgiamo che dal punto di vista di A, ad esempio, le due situazioni sono esattamente identiche ad eccezione del fatto che nella seconda non avviene uno scambio con un'altra persona compresente. Potremmo benissimo ipotizzare infatti che nel secondo caso vengano attribuiti al soggetto gli stessi tratti caratteriali di B, il quale però questa volta rimane a casa comodamente seduto sulla sua poltrona.

Bernard Williams, non mettendo assolutamente in discussione la validità dell'*only x and y principle*, considera le intuizioni che abbiamo riguardo a questi due casi come effettivamente contrastanti, Nozick invece sostiene che la sua Teoria del miglior candidato possa giustificare l'esistenza di intuizioni contrastanti in quanto rende rilevanti nelle due situazioni i fatti estrinseci ad A e B, ed attraverso i quali giustifica l'incompatibilità delle intuizioni. Nel primo caso abbiamo un passaggio delle informazioni dal cervello di A a quello di B e viceversa, di modo che è plausibile definire il processo come uno scambio di corpi. Nella seconda situazione invece, ipotizzando che l'unica differenza con la prima sia il mancato passaggio dei tratti psicologici da A a B e che quindi i tratti psicologici di B vengano di fatto duplicati, secondo la Teoria del miglior candidato di Nozick, alla fine del processo avremo due candidati rivali all'identità con la persona B: l'originale e *closest continuer* e la persona con il corpo di A ed i tratti psicologici di B. La contraddizione messa in evidenza da Bernard Williams viene in questo modo apparentemente risolta, Noonan però fa notare a riguardo che alla fine del processo avremo di fronte anche due *predecessori* della persona A, che ora possiede i tratti di B, in particolare la persona che in partenza possedeva il corpo di A e la persona B. Secondo la *Closest Continuer Theory* di Nozick però, una parte temporale successiva *y* per essere parte dello stesso oggetto continuante di una parte temporale precedente *x*, *y* non deve semplicemente essere il *closest continuer* di *x*, ma *x* deve essere anche il *closest predecessor* di *y*. Dato che i criteri che rendono un individuo il *closest predecessor* di un altro non possono essere diversi da quelli che ne identificano il *closest continuer*, abbiamo però che anche in questo caso quindi i tratti psicologici devono prevalere per importanza su quelli fisici altrimenti non potrebbe descrivere la prima situazione come uno scambio di corpi. In base a questo però la teoria di Nozick non implica che nella seconda situazione A continui a persistere nonostante la radicale trasformazione psicologica e la seguente tortura fisica, ma piuttosto che A cessi di esistere mentre una terza persona C venga alla luce. Dopo il processo di trasferimento infatti A non avrà l'originale A come suo *closest predecessor*, ma bensì B ed al contempo non può essere considerato come il *closest continuer* di B perché questo ruolo rimane saldamente nelle mani di B stesso. Noonan torna a questo punto a criticare la pretesa di Nozick di aver formulato una teoria più potente grazie all'assenza dell'*only x and y principle*, sostenendo che questa conclusione è la stessa alla quale si arriva attraverso la sua teoria dell'identità personale, che pur essendo un Criterialismo psicologico mantiene saldo l'*only x and y principle*, sulla consistenza o meno di una teoria simile tornerò nell'ultima parte di questo testo. Ad ogni modo Noonan afferma che la stessa conclusione si raggiunge semplicemente considerando il fatto che dal momento che nella seconda situazione, dal momento che il grado di continuità psicologica non è sufficiente a stabilire l'identità tra A prima del processo e A al termine dello stesso, l'unico tipo di continuità che si può prendere in considerazione per stabilire l'identità è la continuità fisica, la quale per Noonan non è un criterio

sufficiente. Anche per quanto riguarda la prima situazione descritta le conclusioni cui arriva Noonan sono le stesse a cui arriva Nozick applicando la sua Teoria del miglior candidato in quanto è sufficiente dare un'importanza primaria alla continuità psicologica per poter considerare la prima situazione come uno scambio di corpi.

Secondo Noonan, al contrario, le situazioni descritte da Bernard Williams suggeriscono un argomento contro le Teorie del miglior candidato. Se infatti abbandonassimo la supposizione che il passaggio dei tratti psicologici da B ad A sia stato appositamente causato affidandolo semplicemente al caso e lasciassimo tutto il resto così com'è, avremo il risultato che la persona successiva A non può essere considerata un continuante della persona precedente B, perché la continuità richiede nessi causali per potersi perpetuare, da ciò segue che la persona precedente B non può essere considerata un predecessore della persona successiva A e quindi sicuramente nemmeno un predecessore della persona precedente A. Sembra perciò che in questo caso A, consistentemente con la Teoria del miglior candidato di Nozick, continui ad abitare il proprio corpo nonostante tutti i cambiamenti, mentre la conclusione a cui eravamo arrivati precedentemente era che A cessi di esistere dopo il processo. Questo però, considerato il fatto che i gradi di continuità tra la persona precedente A e la persona successiva A rimangono assolutamente inalterati sia nella situazione originale sia nella variante appena considerata così come gli altri avvenimenti esterni, sembra decisamente assurdo.

L'intento principale di Bernard Williams nel presentare e descrivere queste situazioni era quello di proporre un argomento a sostegno del criterio della continuità fisica per l'identità personale e di evidenziare l'impossibilità di un difensore del criterio psicologico di tracciare una linea di demarcazione tra i punti (i)-(vi), di seguito riportati, per stabilire a che punto diventa fallace sostenere che la persona con il corpo A rimanga A dopo i cambiamenti subiti.

- i. A è soggetto ad un'operazione che produce in lui una totale amnesia.
- ii. L'amnesia viene prodotta in A, ed altre interferenze portano ad alcuni cambiamenti nel carattere di A.
- iii. L'operazione causa i cambiamenti nel carattere di A, ed allo stesso tempo alcune illusorie credenze mnemoniche vengono prodotte in lui; queste comunque sono di carattere fittizio e non rispecchiano quelle di nessuna persona reale.
- iv. Lo stesso che in (iii) eccetto che sia i tratti caratteriali, sia i ricordi sono stati appositamente creati per essere identici a quelli di un'altra persona B.
- v. Lo stesso che in (iv) eccetto che il risultato viene prodotto trasferendo le informazioni dal cervello di B in quello di A, con un procedimento che lascia B identico a come era prima dell'operazione.
- vi. Lo stesso che in (v) eccetto che il procedimento viene svolto anche in senso contrario, in modo che anche le informazioni del cervello di A vengono trasferite nel cervello di B.

Un sostenitore di un criterialismo psicologico per l'identità personale, argomenta Williams, è costretto a dire che dopo i cambiamenti subiti in (vi) la persona con il corpo di A non può più essere definita A ma bensì B e viceversa, allo stesso tempo

però, non può negare che A sopravviva ad i cambiamenti che subisce in (i). A questo punto quindi un ipotetico criterialista psicologico deve essere in grado di tracciare una linea di demarcazione teoretica da qualche parte tra (i) e (vi), ma questo, afferma Williams, non è in grado di farlo. Notiamo che il punto (vi) rappresenta semplicemente la prima situazione descritta da Williams, il punto (v) la seconda e il punto (iv) la variante alla seconda situazione che è stata considerata sopra. Williams però propone il suo argomento dando per scontata la validità dell'*only x and y principle*, un difensore di una Teoria del miglior candidato invece, come è stato già mostrato, è effettivamente in grado di distinguere consistentemente ciò che avviene ad A in ognuno dei casi (i)-(vi) ed è in grado di tracciare due linee di demarcazione, una tra (iv) e (v) e una tra (v) e (vi). Se quindi la *Closest continuer theory* di Nozick è corretta, l'argomento di Williams non è valido. Noonan però sostiene che anche in questo caso non siamo di fronte ad una dimostrazione della maggiore forza esplicativa della teoria di Nozick rispetto alle teorie che mantengono saldo l'*only x and y principle*, in quanto l'argomento di Williams si dimostra, secondo Noonan, ugualmente non valido.

Secondo Noonan, in particolare, l'argomento di Williams dimostra la sua debolezza nel momento in cui si considera il fatto che viene diretto contro un ipotetico sostenitore del criterialismo psicologico disposto a considerare lo scambio delle informazioni tra i due cervelli, mediante la fantastica macchina ipotizzata da Williams, come uno scambio di corpi. Williams però all'inizio del suo saggio, si rivolge esplicitamente alla trattazione di Shoemaker del caso Brown/Brownson e del criterio di identità che ne risulta. In base a questo Noonan strategicamente riformula il punto (v) ed il punto (vi), in modo tale che facciano riferimento allo scambio delle informazioni causato dall'effettivo scambio dei cervelli, ovvero:

- v'. Lo stesso che in (iv) eccetto che il risultato viene prodotto dividendo in due il cervello di B e mettendone metà dentro il cranio di A, dopo aver rimosso il cervello di A, lasciando l'altra metà nel cervello di B.
- vi'. Lo stesso succede ad A come in (v') eccetto che il risultato viene ottenuto scambiando interi i cervelli di A e B mettendoli l'uno nel cranio dell'altro.

A questo punto però, sembra abbastanza facile per un sostenitore del criterialismo psicologico fondato sulla continuità del cervello poter tracciare una linea di demarcazione tra (iv) e (v') sostenendo che fino al punto (iv) viene mantenuta l'identità di A con la persona che occupava il corpo di A prima dell'operazione, mentre oltre no. Bernard Williams nega la possibilità di tracciare una simile linea di demarcazione basandosi su due argomenti. Il primo è che nel passaggio tra (iv) e (v) tutto quello che è stato aggiunto è il fatto che i ricordi e i tratti caratteriali che vengono trasmessi ad A hanno non solo un modello, ma un modello che è anche loro causa, questo però secondo Williams non può fare la differenza dal punto di vista di A, che preoccupandosi del suo futuro avrà sempre paura della tortura annunciata. La seconda ragione in base alla quale Williams nega che possano essere teoricamente distinti i casi (iv) e (v) è che nel caso (v) la persona con il corpo di A non può certamente essere B, in quanto B viene lasciato inalterato e sicuramente queste due non possono essere la stessa persona.

A questo punto però viene chiusa la strada al criterialismo psicologico di sostenere che la persona con il corpo di A in (v) non è più A alla fine del processo perché è

diventata B, ma allora se non è più A, chi è? Se però applichiamo le ragioni di Williams alla serie di casi rivisitata da Noonan, notiamo che entrambi perdono la loro forza argomentativa. La differenza tra (iv) e (v') infatti è che in (v') il cervello di A viene prelevato dal suo cranio e sostituito con metà del cervello di B, di conseguenza la persona che dopo l'operazione abita il corpo di A, è psicologicamente continua con B, di modo tale che i suoi stati mentali saranno causati e dipendenti dalla storia di B, mentre non lo saranno per niente dalla storia di A. In questo modo però difficilmente si potrà sostenere che A sopravviva all'operazione e che non cessi di esistere lasciando posto ad un'altra persona. Se però quest'altra persona non è né A né B allora chi è? Questo ci riporta alla seconda delle ragioni di Williams, nel caso di (v') però non è chiaro perché dovremmo considerare necessariamente la persona nel corpo di A come sicuramente non B in quanto esiste già un indiscutibile continuante di B. Nel caso di (v') infatti la differenza tra la persona nel corpo di A e B è semplicemente il fatto che B mantiene anche la sua continuità fisica. Un sostenitore del criterialismo psicologico può semplicemente considerare questo caso come un caso di fissione, cosa che comporta non pochi problemi, ma comunque in nessuna delle visioni criterialiste riguardo a questo problema uno dei due prodotti della fissione deve essere considerato come indiscutibilmente l'originale. La teoria di Noonan quindi, pur mantenendo saldo l'*only x and y principle*, è riuscita anche in questo caso a rendere consistentemente conto delle nostre intuizioni riguardo a questi casi problematici dell'identità personale mettendo in discussione la maggiore forza esplicativa delle teorie che rinunciano a questo principio.

Anche in relazione ai casi di fissione Nozick sostiene che la sua *Closest continuer theory* renda conto in maniera più convincente ed esaustiva dei problemi da essi derivanti. Stando alla sua teoria dopo una fissione la persona originale cessa di esistere, i casi di fissione, per definizione sono quelli in cui nella gara per la continuità avviene un pareggio, Nozick però afferma che nonostante questo sarebbe assurdo per la persona che sta per subire l'operazione della separazione dei due emisferi, avere per questa operazione la stessa paura che si ha per la morte. Secondo Nozick infatti, nonostante la persona che sta per subire l'operazione stia anche di fatto per cessare di esistere, continueranno ad esistere le due persone che, entrambe, sarebbero identiche a quella persona se solo uno dei due prodotti della fissione sopravvivesse. Sarebbe irrazionale per quella persona pagare qualsiasi cifra per assicurarsi la persistenza di un *closest continuer* e quindi, stando alla teoria di Nozick, la sua persistenza. Nonostante quindi per Nozick l'esistenza di un continuante con un maggior grado di continuità rispetto agli altri non debba essere tenuta in grande considerazione, il futuro di un simile continuante, nel momento in cui se ne ha uno, è una cosa di cui dovremmo razionalmente preoccuparci. Non perché ci sia una proporzionalità tra il grado di continuità di un continuante ed il grado di preoccupazione che la persona originale dovrebbe avere, ma perché nel momento una persona ha un *closest continuer*, indipendentemente dal grado di continuità che ha con lui, razionalmente si preoccupa della sorte e del futuro di chi, secondo Nozick, sarà questa persona. Riconoscendo la contraddittorietà di queste due intuizioni fondate sull'applicazione della sua Teoria del miglior candidato ai casi di fissione Nozick cerca di risolverla appellandosi alla sua interpretazione della struttura dei concetti filosofici e in particolare a quella che chiama la visione della "miglior realizzazione istanziata" di un concetto filosofico. Molto brevemente un concetto ha questa struttura se è possibile descrivere le condizioni di

soddisfazione del predicato espresso in un enunciato della forma: "x soddisfa 'F' se x soddisfa la condizione C, o se niente soddisfa la condizione C, se soddisfa la condizione C'. Soddisfare la condizione C è un modo migliore di istanziare il concetto F, ma se niente soddisfa la condizione C, allora e solo in questo caso anche la condizione C' è sufficiente. È possibile che esista una successione di condizioni C, C', C'', C'''... ognuna preferibile alla successiva, ma ugualmente valida nell'istanziare il concetto solo nel caso in cui niente soddisfi la precedente". Uno dei concetti filosofici che secondo Nozick ha questo tipo di struttura è quello di conoscenza. Uno dei requisiti perché una credenza diventi conoscenza è il fatto che possa variare in base al variare della verità di quanto viene creduto. Il motivo per cui percepiamo le nostre conoscenze come vulnerabili rispetto a posizioni scettiche è il fatto che possiamo immaginare l'esistenza di esseri le cui credenze varino in base ad una applicazione ad un campo molto più vasto di situazioni rispetto al nostro sul quale testano la verità delle loro credenze. Se però, come affermato da Nozick, il concetto di conoscenza ha una struttura del tipo "miglior realizzazione istanziata", il nostro percepire le nostre conoscenze come perennemente vulnerabili non ha ragion d'essere. Questo avviene perché, nel caso in cui i suddetti esseri esistessero realmente allora le nostre credenze semplicemente non sarebbero conoscenze, dal momento in cui però sono semplicemente possibili, le nostre credenze possono qualificarsi come conoscenze in base al variare della loro verità applicata ad un ventaglio di situazioni vasto quanto crediamo che sia. Nonostante la nostra conoscenza non sia il miglior esempio possibile di conoscenza e non sia esempio di conoscenza in altri mondi possibili, in questo mondo si può qualificare come conoscenza.

Ritornando al problema con il quale si era scontrato Nozick, ovvero perché razionalmente noi ci preoccupiamo del nostro *closest continuer*, quando questo esiste, ma non dovremmo preoccuparci della sua unicità nei casi di fissione in cui si hanno più continuanti dello stesso grado? La risposta di Nozick a questa domanda è perché dirigiamo la nostra preoccupazione non verso noi stessi in quanto tali, ma piuttosto verso la migliore realizzazione istanziata di noi stessi. Nel caso in cui io avessi un *closest continuer* quindi, questo sarà la miglior realizzazione istanziata di me stesso e, conseguentemente, le altre realizzazioni istanziate, ovvero gli altri continuanti di minor grado, semplicemente non saranno mie realizzazioni istanziate. Se invece io non disponessi di un *closest continuer*, ma invece di più continuanti dello stesso grado, allora la miglior realizzazione istanziata del concetto di me stesso sarà chiunque di questi soddisfi la condizione di essere un mio continuante di grado non inferiore a nessun altro. In questo modo ognuno dei continuanti di pari grado soddisfa questa condizione e conseguentemente io avrò particolare preoccupazione e cura riguardo al futuro di ognuno di loro, non solo proporzionalmente al grado di continuità che hanno con me ma in quanto realizzazioni del concetto di me stesso. Nonostante l'esistenza di un solo *closest continuer* sarebbe preferibile, in assenza di esso, la compresenza di più continuanti del medesimo grado è sufficiente perché si abbia una realizzazione istanziata del concetto di me stesso e conseguentemente io non avrei motivo di volermi assicurare ad ogni costo la sopravvivenza di un solo *closest continuer*.

Ancora una volta però la soluzione proposta da Nozick non risulta essere molto soddisfacente, Noonan infatti evidenzia immediatamente il fatto che parte del problema, per come era stata inizialmente formulata la questione, era il fatto che mentre il grado di

preoccupazione che abbiamo per il nostro *closest continuer* è indipendente dal grado di continuità che abbiamo con lui, il grado di preoccupazione che abbiamo per tutti gli altri continuanti è direttamente proporzionale al grado di continuità che abbiamo con loro. Nella soluzione che da Nozick però notiamo che il grado di preoccupazione che abbiamo riguardo ai nostri continuanti con lo stesso grado di continuità risultato di episodi di fissione, non è strettamente proporzionale al nostro grado di continuità con loro. Sembra quindi che Nozick abbia cambiato l'oggetto della sua spiegazione nell'arco della stessa, ignorando la sua formulazione originaria.

Anche l'utilizzazione in questo contesto della sua ipotesi riguardo alla struttura dei concetti filosofici è obiettabile. Stando alla sua presentazione originaria della nozione di "miglior realizzazione istanziata di un concetto filosofico" infatti non ha senso la sua applicazione ad un concetto vuoto. Nel momento in cui però la si riferisce al concetto di me stesso, la cui condizione di soddisfazione è semplicemente "un continuante di me stesso che abbia un grado di continuità con me inferiore a quello di nessun altro", il concetto di me stesso viene svuotato, io di fatto non esisto più. Ancora una volta si è rivelata illusoria la pretesa di Nozick e in particolare delle Teorie del miglior candidato, di essere in grado di rendere conto delle intuizioni contrastanti che abbiamo di fronte ai casi enigmatici dell'identità personale come quelli esposti fino ad ora. Il tentativo espresso da Noonan e che verrà esposto nell'ultima parte del testo, è quello di fornire una teoria dell'identità personale che sia in grado di spiegare e descrivere coerentemente le nostre intuizioni fondamentali sull'identità personale mantenendo saldo l'*only x and y principle* e continuando a considerare la continuità psicologica come un criterio sufficiente per l'identità. La plausibilità o meno di quest'ultimo punto in particolare sarà appunto l'argomento dell'ultimo capitolo.

3.5. Gareth Evans e l'obiezione di vaghezza.

Nel suo articolo lungo appena una pagina pubblicato nel 1978 all'interno della rivista *Analysis*, "*Can There Be Vague Objects?*" Gareth Evans considera il fatto che l'affermazione della sufficienza del criterio psicologico per stabilire l'identità personale comporta l'esistenza di casi in cui l'identità personale sia vaga o indeterminata. Se infatti giudichiamo unicamente sulla base della continuità psicologica l'identità di una persona la cui metà dei ricordi e stati mentali sia stata rimossa e rimpiazzata da un'ipotetica macchina in grado di svolgere questo procedimento, risulta indeterminato se questa persona sia o meno la stessa che era in precedenza. L'accezione del termine "vago" utilizzata da Evans è quella applicabile a predicati come "essere calvo". Caratteristica di predicati come questo è infatti quella di non avere confini di applicazione rigidamente definiti. Un uomo al quale si può attribuire il predicato calvo non cesserà di esserlo con l'aggiunta di un singolo capello, l'applicazione di questo predicato avrà di conseguenza una zona "grigia", ovvero ammetterà l'esistenza di casi di confine in cui il valore di verità della frase "*x è calvo*" sarà indeterminatamente vero o falso.

La domanda che si pone Evans nel suo articolo è se la fonte dell'esistenza di enunciati sull'identità indeterminatamente veri o falsi sia linguistica o concettuale, o se

sono gli oggetti stessi che possono essere vaghi. Se consideriamo l'esempio appena riportato della persona i cui stati mentali sono stati per metà sostituiti infatti, notiamo che la soluzione a questo tipo di casi proposta dai sostenitori delle Teorie del miglior candidato è quella di affermare che l'indeterminatezza risieda nel nostro stesso concetto di identità. Esisterebbero quindi per noi, oggetti vaghi, ovvero indeterminatamente identici ad un loro continuante.

La brevità del testo di Evans mi permette di riportarlo per intero:

“ It is sometimes said that the world might itself be vague. Rather than vagueness being a deficiency in our mode of describing the world, it would then be a necessary feature of any true description of it. It is also said that amongst the statements which may not have a determinate truth value as a result of their vagueness are identity statements. Combining these two views we would arrive at the idea that the world might contain certain objects about which it is a fact that they have fuzzy boundaries. But is this idea coherent?

Let ‘a’ and ‘b’ be singular terms such that the sentence ‘a = b’ is of indeterminate truth value, and let us allow for the expression of the idea of indeterminacy by the sentential operator ‘ ∇ ’.

Then we have:

(1) $\nabla(a = b)$.

(1) reports a fact about b which we may express by ascribing to it the property ‘ $x[\nabla(x = a)]$ ’:

(2) $x[\nabla(x = a)]b$.

But we have:

(3) $\sim\nabla(a = a)$

And hence:

(4) $\sim x[\nabla(x = a)]a$.

But by Leibniz’s Law, we may derive from (2) and (4):

(5) $\sim(a = b)$

Contradicting the assumption, with which we began, that the identity statement ($a = b$) is indeterminate in truth value.

If ‘Indefinitely’ and its dual ‘Definitely’ (Δ) generate a modal logic as strong as S5, (1) - (4) and, presumably, Leibniz’s Law, may each be strengthened with a ‘Definitely’ prefix, enabling us to derive

(5') $\Delta\sim(a = b)$

which is straightforwardly inconsistent with (1)."

(Can There Be Vague Objects? in Analysis 38.4, p. 208. 1978)

Come viene fatto notare da Garrett (1998) nell'esposizione che propone di questa dimostrazione, nella prima parte dell'articolo Evans distingue tra due visioni sulla vaghezza: la visione secondo la quale è il mondo stesso ad essere vago e quella secondo

cui la vaghezza è il risultato di una deficienza nel nostro modo di descrivere il mondo. Il passaggio fondamentale per capire la prova di Evans risiede nel suo constatare che se è il mondo stesso ad essere vago, allora devono esistere enunciati di identità vaghi, la cui vaghezza non risiede nell'indeterminatezza referenziale dei termini singolari usati, ovvero non risiede in una deficienza del nostro modo di descrivere e di riferirci al mondo. Quello che stabilisce quindi con la sua dimostrazione è appunto che non possono esistere enunciati di questo tipo. La dimostrazione di Evans è consistente con la possibilità che esistano vaghi enunciati di identità. Questo è un fatto innegabile, se infatti ipotizziamo l'esistenza di una fila di penne i cui colori procedano impercettibilmente e adiacentemente dall'arancione al rosso, la descrizione definita "la prima penna rossa" non ha un referente determinato. Conseguentemente l'affermazione "la mia penna è la prima penna rossa" sarà un enunciato di identità vago. Quello che Evans nel suo articolo tenta di stabilire è che la vaghezza di questo enunciato deriva dall'indeterminatezza referenziale della descrizione definita e non dalla vaghezza del colore della penna. La prova di Evans è quindi consistente con la vaghezza della frase "la mia penna è la prima penna rossa", quello che viene negato è il fatto che possa esistere un enunciato di identità vago $\nabla(a = b)$ tale che implichi $(\exists x)(\exists y) \nabla(x = y)$. Questa infatti sarebbe una prova esplicita dell'indeterminatezza dell'identità. L'esempio sopra riportato riguardo al vago enunciato di identità che si riferiva alla prima penna rossa, non implica invece l'indeterminatezza dell'identità in quanto il termine singolare "la prima penna rossa" non individua determinatamente una tra le penne affermando che questa è indeterminatamente la mia penna. Ancora una volta, l'indeterminatezza non risiede nell'oggetto ma nel nostro modo di individuarlo.

La prova di Evans ha la forma di una *reductio ad absurdum* dalla struttura molto chiara. In particolare (1) implica (2) e (3) implica (4), (2) e (4) per la Legge di Leibniz implicano (5), (1) - (5) infine possono essere rafforzati attraverso l'uso del prefisso Δ , ottenendo così (5'), il quale è in contraddizione evidente con (1). Come è stato reso evidente dall'esempio sulla prima penna rossa però, perché la prova di Evans funzioni correttamente, è necessario restringere il campo dei possibili a e b a quello delle costanti logiche che individuano oggetti senza contenuto descrittivo. Esaminando la sua dimostrazione ad ogni modo è possibile notare che (2) è una tautologia ed è quindi lecito chiedersi perché Evans non sia partito direttamente da questo punto per il suo ragionamento. Le ragioni per cui ha scelto (1) come punto di partenza per il suo ragionamento sono due. La prima consiste nel voler evitare l'analogia con il procedimento fallace che porta da un'affermazione *de dicto* come (i): "John crede che la spia più alta sia una spia." ad una *de re* come (ii): "la spia più alta è tale che John crede che sia una spia". L'enunciato (i) potrebbe infatti essere vero semplicemente perché John suppone che esista, senza conoscerla direttamente, una spia più alta, ed è conscio del fatto che "la spia più alta è una spia" è una tautologia. In questo caso però l'enunciato (ii) sarebbe falso perché richiede che John conosca direttamente l'uomo che è la spia più alta. Similmente si potrebbe essere arrivati a (2) attraverso un procedimento non valido, motivo per cui Evans ha ritenuto necessario esplicitare il passaggio. La seconda e più importante ragione della scelta di Evans consiste nel fatto che è in (2) piuttosto che in (1) che si caratterizza maggiormente l'idea che l'identità, intesa come individuo extra-linguistico, possa avere forma indeterminata. Arrivando formalmente e indiscutibilmente a (2) Evans intende quindi rafforzare la posizione della

visione che sta cercando di controbattere. Se (2) fosse stato attaccabile infatti, l'intera dimostrazione, che è necessario ricordare ha forma di *reductio ad absurdum*, perderebbe la sua validità, Evans avrebbe vinto in partenza perché non ci sarebbe stata nessuna posizione stabile da attaccare.

Garrett individua tre questioni, tre punti sensibili facendo leva sui quali la dimostrazione di Evans può essere attaccata. La prima domanda che pone Garrett riguardo alla prova di Evans è se $x[\nabla(x = a)]$ denoti realmente una proprietà. Questa richiesta è essenziale perché la dimostrazione sia valida; se infatti la formula sopra riportata non denotasse una proprietà nel corretto senso del termine la prova di Evans sarebbe analoga al sillogismo fallace espresso attraverso l'argomento del "... è così chiamato a causa della sua stazza". Questo argomento è rappresentato dall'inferenza:

Giorgione era chiamato così a causa della sua stazza.
Barbarelli non era chiamato così a causa della sua stazza; quindi
Giorgione non è Barbarelli.

Questa inferenza è invalida a causa del fatto che il predicato: "è chiamato così a causa della sua stazza" non denota una reale proprietà dell'individuo a cui si riferisce in quanto la verità o la falsità della sua ascrizione ad un individuo dipende dal *come* ci stiamo riferendo a quell'individuo. In questo caso se ci stiamo riferendo a lui come Giorgione o come Barbarelli.

Se quindi $x[\nabla(x = a)]$ non denotasse una reale proprietà dell'oggetto a cui si riferisce la prova di Evans sarebbe analogamente invalida, ma perché non ci sarebbe semplicemente motivo di ritenere possibile che l'identità possa essere vaga o indeterminata, Evans non avrebbe semplicemente avuto bisogno di formulare la sua dimostrazione. Allo scopo di determinare se la prova di Evans arrivi correttamente alla stessa conclusione anche considerando l'indeterminatezza dell'identità di un oggetto con un altro come una sua proprietà genuina, si assuma momentaneamente che ciò sia vero in modo da poter procedere nella discussione.

La seconda questione che Garrett pone alla nostra attenzione è il fatto che la validità della prova di Evans può essere messa in discussione in due punti. In particolare nell'inferenza che porta da (2) e (4) a (5) e in quella che porta da (5) a (5'). Per quanto riguarda la prima delle due inferenze si potrebbe obiettare ad Evans che $x[\nabla(x = a)]$ denoti una proprietà diversa in (2) rispetto a quella che denota in (4). Questo equivarrebbe a sostenere che il referente del predicato varia quando questo viene applicato ad un soggetto diverso. Se applicassimo questa obiezione all'esempio di Giorgione appena riportato, si otterrebbe che il predicato "... viene chiamato così a causa della sua stazza" esprimerebbe la proprietà "essere chiamato Giorgione a causa della sua stazza" se applicato al soggetto "Giorgione", mentre esprimerebbe la proprietà "essere chiamato Barbarelli a causa della sua stazza" se applicato a "Barbarelli". Mentre però questa ambiguità spiega la confusione alla radice dell'apparente validità dell'inferenza che porta alla conclusione dell'argomento su Giorgione, non esiste un reale motivo per sostenere che il referente dell'espressione $x[\nabla(x = a)]$ sia determinato dal termine al quale viene associato.

Il passaggio da (5) a (5') richiede invece una maggiore attenzione. Nell'ultimo paragrafo della sua dimostrazione Evans inserisce il termine "definito" ed il relativo

simbolo Δ come duale e contrario di “indefinito”, sostenendo che se in questo modo viene a formarsi una logica modale forte quanto quella definita da (1) - (4) e dalla legge di Leibniz allora è possibile rafforzare ciascuna formula con il prefisso Δ . L'affermazione di Evans che indica come duali i termini “definito” e “indefinito” però, è vera soltanto se “ Δ ” viene letto come non fattivo, ovvero solo se Δp non implica p . Il principio in grado di giustificare la plausibilità del rafforzamento di (2) e (4) con il prefisso Δ , è quello per il quale $\nabla p \rightarrow \Delta(\nabla p)$, questo principio però risulta essere non valido se considerato in rispetto a casi di vaghezza di ordine superiore. Nel caso in cui infatti, ∇p sia esso stesso indeterminato, un caso cioè di vaghezza di secondo ordine, il condizionale $\nabla p \rightarrow \Delta(\nabla p)$ avrebbe un antecedente indeterminato ed il suo conseguente dovrebbe quindi essere considerato falso. Un condizionale con un antecedente indeterminato ed un conseguente falso però, plausibilmente non sarà vero ma piuttosto indeterminato e conseguentemente non potrà avere la funzione di assioma di una logica modale. Sembra quindi che nessuna possibile logica della vaghezza possa essere forte quanto il sistema modale definito in (1) - (5). Nonostante però $\langle \nabla, \Delta \rangle$ non generino una logica modale forte quanto quella espressa nella seconda parte dell'articolo, la dimostrazione di Evans sembra ugualmente mantenere la sua validità riguardo ai casi fondamentali e centrali cari ai difensori della possibile indeterminatezza dell'identità. Quanto presumibilmente viene sostenuto da un difensore di questa visione è infatti che almeno alcune vaghe relazioni di identità sono definitivamente vaghe. Nel caso ad esempio della sequenza di penne che procedono dal rosso all'arancione un sostenitore della possibile vaghezza dell'identità affermerà plausibilmente che le penne intorno alla metà della sequenza sono esempi definiti della zona grigia e indefinita tra rosso e arancione. Riguardo a casi come questi quindi, Evans può consistentemente aggiungere il prefisso Δ a (2) e (4) per ottenere (5).

Come viene fatto notare nella trattazione della prova fornita da Garrett ad ogni modo, non è completamente chiaro il motivo per cui Evans abbia voluto aggiungere l'ultima parte, (5) infatti contraddice (2) in maniera assolutamente inquestionabile in quanto (5) è vero se e solo se $a = b$ è falso e la falsità di $a = b$ è incompatibile con (2).

La terza ed ultima prova a cui Garrett sottopone la dimostrazione di Evans è se la premessa (4) $\sim x[\nabla(x = a)]a$, ovvero l'assunto che a non possa avere la proprietà di essere tale che sia indeterminato se sia o meno identico ad a , possa indubitabilmente essere considerata vera. Nonostante questa sembri una supposizione più che plausibile in quanto nel momento in cui abbiamo individuato un oggetto sembra insensato chiedersi se questo sia solo indeterminatamente identico a se stesso, a questa affermazione potrebbe essere obiettato che in realtà si stanno confondendo due tesi diverse. Una tesi è quella che considera a come definitivamente identico a se stesso, mentre l'altra è quella che considera a come definitivamente identico ad a . Quella necessaria per la validità della prova di Evans è la seconda tesi ma potrebbe essere obiettato che la verità della prima non implica la verità della seconda. Anche se questo fosse vero comunque, non sussiste ugualmente alcuna ragione per mettere in questione la verità di (4) che continua ad essere un assunto valido di per sé.

Un ipotetico sostenitore della possibile vaghezza degli oggetti però potrebbe affermare che la verità di (4) è esattamente quanto la sua ipotesi intende negare. Quello che potrebbe essere obiettato alla prova di Evans è il fatto che per “oggetto vago” si intende precisamente un oggetto x tale che sia indeterminato se è identico o meno a x .

Se le cose stessero così allora la prova di Evans, il cui intento è quello di escludere la possibilità dell'esistenza di oggetti vaghi, scanserebbe semplicemente il problema perché assumerebbe ad assioma ciò che intende affermare. La risposta che è stata data da David Wiggins a questa obiezione è che, anche ammettendo che *a* sia un oggetto vago, dovremmo comunque essere in grado di stabilire un perfetto caso di identità associando *a* con se stesso, in quanto sussisterebbe ugualmente una perfetta corrispondenza anche relativamente alla sua vaghezza. Il pericolo insito in questa risposta risiede nel fatto che se si considera (2) consistente con la caratterizzazione di *a* come oggetto vago, allora anche la formula derivabile da (5'): $\sim x[\nabla(x = a)]b$ deve essere consistente con la possibilità che *b* sia un oggetto vago, ma in questo caso la prova di Evans non potrebbe essere usata per affermare l'impossibilità dell'esistenza di oggetti vaghi. Sembra quindi, che si sia arrivati ad una impasse, in particolare o (2) è inconsistente con la possibilità che *a* sia un oggetto vago e quindi la prova di Evans assume per dato ciò che intende affermare, oppure anche l'ultima formula sopra riportata è consistente con la possibilità che *b* sia un oggetto vago, e quindi la prova di Evans fallisce nel suo intento.

La soluzione proposta da Garrett per uscire da questo vicolo cieco consiste nel riconsiderare quanto viene scritto da Evans nel primo paragrafo dell'articolo, ovvero l'affermazione dell'incoerenza delle idee della possibilità dell'esistenza di oggetti vaghi e della determinatezza della relazione di identità. In particolare il condizionale alla base della prova di Evans è: se lo stabilirsi o meno della relazione di identità tra due oggetti è sempre un fatto determinato, allora non possono esistere oggetti vaghi. Mentre da un certo punto di vista questo condizionale è perfettamente giustificato da un altro notiamo però che esistono altri criteri per l'esistenza di oggetti vaghi che sono consistenti con la determinatezza della relazione di identità. Questi criteri sono quelli di natura referenziale, che Evans non ha mai inteso negare, come ad esempio il fatto che un oggetto venga considerato vago nel caso in cui sia vago il suo possedere o meno una determinata parte. Nonostante quindi la formula derivata da (5') sia consistente con l'esistenza di oggetti vaghi, la prova di Evans mantiene comunque la sua validità nello stabilire che la relazione di identità non può essere vaga o indeterminata, ma che tale vaghezza risiede esclusivamente nel nostro modo di riferirci agli oggetti.

Applicata all'identità personale quindi, la prova di Evans stabilisce che questa non possa essere indeterminata, ma che al contrario, l'indeterminatezza di certi enunciati sull'identità personale è di natura referenziale. Questo la rende ugualmente consistente con le conseguenze dell'argomentazione di Bernard Williams, ovvero il fatto che l'innalzamento di un criterio a necessario e sufficiente per l'identità personale, comporta lo stabilirsi di vaghi enunciati di identità. Stabilisce formalmente però, che la vaghezza e indeterminatezza di questi enunciati è di natura esclusivamente linguistica e non ontologica come viene proposto dai sostenitori delle Teorie del miglior candidato. La relazione di identità non può essere indeterminata.

3.6. Conclusioni.

In questo capitolo ho cercato di mostrare come le critiche di Noonan alle teorie di

Parfit e di Nozick colpiscano il segno e mettano in evidenza quella che credo che sia la loro contraddizione di fondo, ovvero il fatto che arrivino a soluzioni radicalmente controintuitive per poter rendere conto delle nostre intuizioni riguardo ai casi enigmatici sull'identità personale. Ho cercato inoltre di criticare come ingiustificabile alla luce delle argomentazioni sulla relazione di identità di Wiggins e della consistenza della prova di Evans, il loro rifiutare l'*only x and y principle* ed adoperare un concetto di identità che la vede passibile di essere vaga o indeterminata. Nel prossimo capitolo verrà presentata la proposta teorica di Noonan, il quale tenta di delineare un criterialismo psicologico basato su una concezione dell'identità che ha tra i suoi principi fondamentali l'*only x and y principle* e che sappia ugualmente descrivere efficacemente le nostre intuizioni riguardo ai casi di fissione, teletrasporto e sull'identità personale in generale. Anche la formulazione di Noonan però non è esente da problemi alcuni dei quali vengono da lui stesso riconosciuti nell'ultima parte di *Personal Identity* (2003). Credo inoltre, che sia obiettabile l'argomentazione attraverso la quale Noonan supera l'obiezione che Bernard Williams rivolge ai sostenitori di un criterialismo psicologico e che, considerando il fatto che l'obiezione di Williams se vista più attentamente si rivolge ai criterialisti in generale, la soluzione proposta da Noonan abbia il risultato di chiuderlo in un vicolo cieco.

4.

LA PROPOSTA DI NOONAN E LE SUE DIFFICOLTA'.

In quest'ultimo capitolo verrà presa in considerazione la proposta teorica di Harold Noonan, il quale ha tentato di associare consistentemente un criterialismo psicologico sull'identità personale ed una visione dell'identità che la caratterizza come assoluta e determinata. Per quanto sue le argomentazioni siano molto interessanti, Noonan stesso sia in *Personal Identity* (2003), sia nell'articolo *Personal Identity And Its Perplexities* (2012), dimostra di essere cosciente di alcuni punti deboli della sua formulazione. Personalmente inoltre, trovo che sia criticabile il modo in cui Noonan, nel decimo capitolo di *Personal Identity* (2003), supera il *Reduplication Argument* di Bernard Williams; passaggio fondamentale per poter arrivare a proporre la sua originale teoria.

La tesi che verrà argomentata in questo capitolo è quindi quella che sostiene che Noonan, nel tentativo di associare consistentemente criterialismo psicologico e *only x and y principle*, sia stato costretto a scoprire il fianco ad altre obiezioni ugualmente cogenti. Verrà mostrato infatti, come Noonan per poter superare le obiezioni di Bernard Williams, sia costretto ad operare una rivoluzione in campo ontologico mettendo in discussione la nostra tradizionale concezione che vede gli oggetti come persistenti nel tempo. Noonan adotta infatti per la sua teoria una concezione ontologica quadrimensionalista che presenta di per sé non pochi problemi e che a mio modo di vedere, è altrettanto controintuitiva di una concezione dell'identità che fa a meno dell'*only x and y principle*.

4.1. In difesa del Reduplication Argument.

Nel capitolo dieci di *Personal Identity* (2003) Noonan rivolge alcune critiche, che sono già state presentate nel capitolo precedente, al *Reduplication Argument* proposto da Bernard Williams in *The Self And The Future* (1970). Allo scopo di mostrare come le obiezioni stesse formulate da Noonan non siano esenti da possibili critiche, sarà necessario riproporre brevemente dal principio i passaggi argomentativi percorsi da Noonan.

Tenendo a mente che l'argomento di Williams si basa sulla considerazione di una situazione che lui definisce come quella di uno scambio di corpi tra due soggetti A e B e che, in ognuno dei casi sotto riportati, si suppone che successivamente a quanto

descritto la persona con il corpo di A venga torturata e che prima del trattamento A venga informato di quanto sta per succedergli e della successiva tortura. Williams propone la seguente successione di casi.

- i. A subisce un'operazione che produce in lui una totale amnesia.
- ii. Viene prodotta l'amnesia in A, e altre interferenze portano ad alcuni cambiamenti nel suo carattere.
- iii. Vengono prodotti dei cambiamenti nel carattere di A, e allo stesso tempo vengono introdotti in lui alcune credenze mnemoniche illusorie: queste hanno carattere fittizio e non rispecchiano quelle di un'altra persona.
- iv. Lo stesso che in (iii), eccetto che sia i tratti caratteriali, sia i ricordi sono stati disegnati in modo tale da rispecchiare quelli di un'altra persona.
- v. Lo stesso che in (iv), eccetto che il risultato viene prodotto trasferendo le informazioni dal cervello di B a quello di A, attraverso un metodo che lascia B immutato.
- vi. Lo stesso succede ad A come in (v), mentre B non viene lasciato immutato, in quanto un'operazione simile viene svolta anche nella direzione opposta.

Quanto viene sostenuto da Bernard Williams è che nei primi tre casi evidentemente non si ha alcuna ragione per pensare che A non dovrebbe temere la futura prospettiva di essere torturato. Anche per quanto riguarda il quarto caso inoltre, nonostante venga introdotta una seconda persona B, non sembra che questa venga introdotta in un modo tale che possa avere qualche ripercussione sui timori che A dovrebbe avere riguardo al suo futuro. L'unica differenza tra (iii) e (iv) risiede nel fatto che i ricordi che vengono inseriti nel cervello di A hanno un modello. Nonostante le cose sembrino complicarsi quando consideriamo il caso (v), l'opinione di Williams a riguardo è che di fatto l'unica differenza con il punto (iv) è che in (v) il modello B dei ricordi e dei tratti caratteriali che vengono trasmessi ad A è anche loro causa, ma questo continua a non avere effetti sul fatto che A dovrebbe essere spaventato al pensiero che il suo corpo dopo il procedimento verrà torturato. Se infatti A è in grado di aspettarsi del dolore, è anche in grado di aspettarsi un dolore successivo ad un cambiamento nelle sue disposizioni, il fatto che questo cambiamento nelle sue disposizioni abbia un modello di riferimento o che sia indirettamente causato dalle disposizioni di un'altra persona non fa nessuna differenza. Nella situazione (v) inoltre, è ancora fuori discussione l'argomento che vede l'A post-operazione identificarsi con B in quanto B rimane invariato dopo l'operazione e rimane quindi anche indiscutibilmente la stessa persona, di modo tale che non si possa dire che A sia diventato B. La situazione (vi) però, apparentemente sembra passibile di essere definita come uno "scambio di corpi", ovvero una situazione in cui la persona A si trova ora nel corpo di B e viceversa. Se però nella situazione (v) le paure e le preoccupazioni di A per il suo futuro possono estendersi oltre il cambiamento delle sue disposizioni, così come avviene in (iii) e (iv), allora lo stesso dovrebbe avvenire anche nella situazione (vi), dato che dal punto di vista di A nulla è cambiato. Relativamente a quanto accade ad A infatti, non c'è assolutamente nessuna differenza tra quanto avviene in (v) e quanto avviene in (vi), se quindi razionalmente avrebbe dovuto temere per il suo futuro in (v), lo stesso dovrebbe fare in (vi).

Sostenere che tra (v) e (vi) sussista una differenza sostanziale che si rispecchia in un

cambiamento razionalmente fondato nelle aspettative e paure di A, significa negare che l'unica differenza tra (v) e (vi) riguardi quanto succede a B e sostenere invece che, diversamente da quanto accade in (v), la situazione (vi) implichi la reintroduzione della persona A, nel corpo di B diventato il portatore dei suoi ricordi e tratti caratteriali. Secondo questa interpretazione della situazione esposta è per il futuro di questa persona, che risiede nel corpo di B, che A dovrebbe razionalmente avere aspettative e preoccupazioni. Una conseguenza di questa visione dell'identità personale però è il fatto che allora nella situazione (v) A non dovrebbe preoccuparsi del futuro della persona che risiederà nel suo corpo. Se infatti in quest'ottica la persona con il corpo di A non è A ma B in quanto possiede tutti i suoi ricordi e tratti caratteriali e viceversa, A dovrebbe arrivare alle stesse conclusioni riguardo a se stesso anche nella situazione (v), in quanto nulla è cambiato riguardo ai ricordi ed ai tratti caratteriali che abitano nel corpo di A dopo l'operazione. Se la persona che abita il corpo di A nella situazione (vi) dopo l'operazione non è più A, allora lo stesso deve accadere nella situazione (v), nonostante in (v) non esista nessuna persona che possa razionalmente candidarsi al titolo di identica con quella che abita il corpo di A dopo l'operazione, A semplicemente smetterebbe di esistere.

Anche ammettendo momentaneamente questa conclusione, rimane a questo punto difficile tracciare una linea di demarcazione tra (v) e le precedenti quattro situazioni esplicitando fino a quale tra queste A sopravviva all'operazione. Una possibile risposta è quella di ammettere casi limite in cui è concettualmente indecidibile se la persona risultante dall'operazione sia la stessa che era in precedenza o meno. L'argomento di Williams però, anche ammettendo l'indecidibilità, non l'indeterminatezza, di alcuni casi, non perde la sua validità, esistono infatti altre emozioni vicine od alternative alla paura, che insorgono similamente a questa quando ci confrontiamo con situazioni il cui esito è incerto. È sufficiente quindi chiarificare di volta in volta quali emozioni siano razionalmente fondate ed in quale misura perché l'argomentazione di Williams mantenga la sua validità, fermo restando che, a processo avvenuto, la persona risultante dall'operazione è determinatamente la stessa che era precedentemente o meno.

Un'altro aspetto dell'argomento di Williams che secondo lo stesso autore è importante sottolineare, è il fatto che sembra contraddire quanto spesso si suppone più o meno velatamente riguardo all'identità personale. Nello specifico Williams si riferisce al fatto che spesso si distingue tra aspetti "in prima persona" e aspetti "in terza persona" dell'identità personale e che molte difficoltà risiedano nella relazione tra queste due categorie di aspetti. Spesso in particolare considerazioni di carattere psicologico vengono associate a chi propone un approccio che si focalizza sugli aspetti "in prima persona" dell'identità personale. Mentre chi difende un approccio che privilegia gli aspetti "in terza persona" si concentra su considerazioni fisicaliste e sulla continuità del corpo. Quello che Williams sostiene di aver mostrato attraverso la sua argomentazione è che in realtà avviene esattamente il contrario, mentre infatti la prima parte dell'argomento di Williams, mi sto in questo momento riferendo alla formulazione originale del *Reduplication Argument* esposta nel terzo capitolo, ovvero quella che porta a conclusioni di tipo psicologista, è costruita interamente "in terza persona", la seconda parte del suo argomento, quella che conduce invece a conclusioni fisicaliste, è basata sulle aspettative e paure che dovrebbe "in prima persona" avere A nelle varie situazioni.

Secondo Williams le considerazioni che portavano nella prima delle situazioni

descritte a identificare il processo con quello di uno “scambio di corpi” erano solo artificialmente fondate in base al volere dello sperimentatore di creare una situazione che si avvicinasse il più nettamente possibile a questo concetto scegliendo attentamente come svolgere l’operazione. L’unico modo che abbiamo, secondo Williams, di considerare l’esperimento come un effettivo scambio di corpi, è quello di concepire le persone come dei fantasmi che abitano i corpi e che possono essere spostati e scambiati attraverso una qualche procedura. Williams si sta riferendo al fatto che se consideriamo la continuità psicologica come un criterio sufficiente per l’identità personale allora abbiamo simultaneamente identificato sia il “fantasma” della persona che abita un corpo, sia il metodo di trasmissione della persona da un corpo ad un altro, ovvero qualsiasi metodo che preservi la continuità psicologica della persona. Ad ogni modo secondo Williams questa visione dell’identità personale è semplicemente fallace. Lo sperimentatore nei casi sopra citati non ha prodotto un effettivo scambio di corpi, ma ha semplicemente portato avanti quello, tra i processi possibili, che saremmo stati maggiormente disposti a chiamare uno scambio di corpi.

Come è stato già visto nel terzo capitolo la soluzione proposta da Noonan per superare l’argomento di Williams, consiste nel restringere il campo del *Reduplication Argument* relativizzandolo le sue argomentazioni in modo che si applichino esclusivamente ad un criterialismo psicologico fondato sulla continuità del cervello. In particolare sostituisce ad i punti (v) e (vi) come erano stati originariamente formulati i punti (v’) e (vi’) da lui appositamente modificati per lo scopo.

- v'. Avviene lo stesso come in (iv), eccetto che il risultato viene prodotto separando gli emisferi del cervello di B e mettendone uno dentro il cranio di A, dopo aver rimosso il cervello di A, lasciando l’altro emisfero nel cranio di B.
- vi'. Lo stesso accade ad A come in (v’), eccetto che il risultato viene prodotto traferendo l’intero cervello di B nel cranio di A e viceversa.

Sostituendo agli originali punti (v) e (vi) le due situazioni sopra riportate Noonan afferma di essere in grado consistentemente di tracciare una linea di demarcazione tra il punto (iv) ed il punto (v’), sostenendo che fino al punto (iv) viene mantenuta l’identità del soggetto sottoposto all’esperimento, mentre dalla situazione (v’) in poi questa viene meno.

Nonostante attraverso questa rivisitazione ad hoc dei casi proposti da Williams Noonan sembri in grado di riuscire a superare tutte le questioni che venivano considerate *Reduplication Argument*, rimane a mio parere almeno un’argomentazione posta da Williams che può a sua volta essere relativizzata per essere rivolta contro una versione del criterialismo psicologico fondato sulla continuità del cervello. In particolare la serie di casi rivisitata da Noonan ha senso solo se messa in relazione con un criterialismo psicologico che sostiene che quando avviene una fissione, l’identità con il soggetto di partenza non viene mantenuta da nessuno dei prodotti della fissione. Se infatti si accetta il fatto che, dato un caso di fissione come quello presentato nella situazione (v’), uno dei due prodotti della fissione mantiene l’identità con il soggetto di partenza, in questo caso verosimilmente basandosi sulla continuità del corpo, cosa di per sé già discutibile per una teoria che sostiene che la continuità psicologica sia un

criterio sufficiente per l'identità personale, allora sarebbe il prodotto della fissione con il corpo di B a mantenere l'identità con B, e la soluzione proposta da Noonan attraverso la serie rivisitata di casi si rivelerebbe fallace. Questo avviene perché una teoria simile non riuscirebbe a spiegare come mai in (v') B permanga come identico e A cessi di esistere senza diventare B, mentre in (vi') la persona con il corpo di A sia identica a B e viceversa. Se infatti B permane come identico sia che resti con un solo emisfero nel suo corpo, sia che li mantenga entrambi nel corpo di A, non si spiega come mai in un caso la continuità del corpo venga presa in considerazione per stabilire l'identità di uno dei due prodotti della fissione con il soggetto di partenza mentre nell'altro no, negando in (v) l'identità con B alla persona con il corpo di A e rendendola un semplice continuante. Mentre per chi considera *l'only x and y principle* valido questo è un problema reale, un sostenitore di una Teoria del miglior candidato non avrebbe alcun problema a stabilire una classifica di criteri per l'identità personale in cui l'identità del cervello sia più importante dell'identità del corpo, ed a sostenere che nel caso in cui non si abbiano altri candidati in cui viene mantenuta l'una, allora si deve prendere in considerazione l'altra per stabilire l'identità, i problemi che questa tipologia di teorie comporta però, sono già stati affrontati.

Per fare in modo che la successione dei casi proposta da Noonan per superare le difficoltà esposte da Bernard Williams svolga efficacemente il suo compito, rimangono quindi aperte due possibilità: considerare entrambi i prodotti della fissione come identici all'originale oppure non considerarne nessuno. La prima di queste due possibilità è evidentemente assurda in quanto considera numericamente identiche due entità simultaneamente occorrenti. La seconda possibilità è a sua volta passibile di due interpretazioni diverse: è possibile infatti considerare i prodotti della fissione come persone nuove che cominciano la loro esistenza nel momento stesso in cui avviene la fissione, oppure, come propone Noonan, come persone già esistenti precedentemente alla fissione. Dal momento che la categoria di teorie che Bernard Williams intende attaccare con il *Reduplication Argument* è quella dei criterialismi psicologici, ristretta dalla serie rivisitata dei casi (i) – (vi') proposta da Noonan a quella dei criterialismi psicologici in cui la continuità mnemonica è fondata sulla continuità del cervello, la difesa della prima delle due interpretazioni implica l'adozione di una teoria dell'identità personale che fa a meno dell'*only x and y principle*, ovvero di una Teoria del miglior candidato. L'unico motivo per cui l'originale non può essere identificato né con A né con B in quest'ottica è il fatto che una cosa non può essere numericamente identica a due successive e contemporaneamente esistenti. Secondo questa visione dei casi di fissione uno dei due continuanti dopo l'operazione può veridicamente asserire una frase come: "fortunatamente l'altro emisfero non è stato distrutto altrimenti io non sarei esistito perché sarei stato identico all'originale". Le conseguenze a mio parere assurde alle quali si arriva utilizzando un concetto di identità che determina l'identificazione di un oggetto con un altro successivo anche in base a fattori estrinseci agli oggetti stessi però sono già state mostrate nell'arco del terzo capitolo.

L'unica possibilità rimasta ad un sostenitore di un criterialismo psicologico fondato sulla continuità del cervello per poter tracciare consistentemente una linea di demarcazione tra i casi (iv) e (v'), è quella che viene effettivamente percorsa da Noonan nel dodicesimo capitolo di *Personal Identity* (2003). Il prezzo però che Noonan deve pagare per poter ammettere la preesistenza all'operazione di ognuno dei continuanti

risultanti dalla fissione, è il fatto che si trova costretto a dover adottare un'ontologia di tipo quadrimensionalista che aggiunge alle tre dimensioni dello spazio tradizionali, una quarta dimensione temporale. Le persone, stando a questa visione ontologica, sono composte di *person-stages* che si succedono nel tempo. Stando così le cose le persone, così come gli altri oggetti, non sarebbero completamente presenti in ogni momento della loro esistenza come viene sostenuto dai difensori della più tradizionale ontologia tridimensionalista. In questo modo Noonan cerca di rendere conto della compresenza spaziale di più persone nello stesso corpo precedentemente alla fissione ammessa dalla sua teoria, la quale verrà più esaurientemente esposta nella prossima parte. Apparentemente Noonan riesce in questo modo a giustificare le sue affermazioni mantenendo saldo l'*only x and y principle*, ma se viene considerata più attentamente anche la sua formulazione non è esente da problemi, sia di per sé, sia derivanti dall'aver accettato come valida un'ontologia di tipo quadrimensionalista.

Oggetto della parte successiva di questo capitolo sarà presentare la proposta teorica di Noonan e di metterne in risalto i punti deboli. É stato precedentemente argomentato che quella percorsa da Noonan inoltre era l'unica strada rimasta tra quelle utilizzabili da un sostenitore di un criterialismo psicologico, anche se fondato sulla continuità del cervello, per superare le obiezioni mosse a queste teorie da Bernard Williams. Nel momento in cui anche questa via si rivela problematica ed attaccabile però il *Reduplication Argument* dimostra ancora una volta la sua forza argomentativa e la sua attualità di ostacolo difficilmente superabile per chi sostiene che la continuità psicologica, sia essa fondata sulla continuità del cervello o meno, sia un criterio sufficiente per l'identità personale.

4.2. La teoria di Noonan.

Dopo aver criticato la pretesa delle Teorie del miglior candidato di avere una maggiore potenza esplicativa ed aver mostrato come in realtà la teoria di Nozick non sia così conforme alle nostre intuizioni sull'identità personale come il suo autore voleva sostenere, Noonan nell'ultima parte di *Personal Identity* (2003) delinea brevemente la sua originale proposta teorica.

La differenza fondamentale tra la formulazione di Noonan e quella di Nozick risiede nel fatto che Noonan considera ogni linea di continuità sufficientemente forte come la storia di un qualche individuo, indipendentemente da quanto accade all'infuori di lui. Ciò implica che entrambi i prodotti di una fissione debbano esistere anche precedentemente alla stessa. La teoria di Nozick al contrario, facendo a meno dell'*only x and y principle*, nega questa conseguenza. La soluzione che viene proposta da Noonan implica che ognuna delle due *persone* in esistenza successivamente ad una fissione, siano esistite indipendentemente anche precedentemente all'operazione e questa visione ha delle conseguenze che si riflettono sul concetto stesso di *persona* che viene in questo modo delineato. Per fare in modo di disporre di un concetto di persona che sia compatibile con la soluzione da lui proposta per i casi di fissione, Noonan suddivide la persona in *person-stages* temporali, individuando all'interno di questa successione un

membro, da lui chiamato “marker” tale che tutti e solo i membri della successione siano suoi migliori continuanti o predecessori e chiama successivamente questa successione un *person-set*. Noonan fa quindi coincidere l'estensione del termine “persona” con un *person-set*, mentre l'unione di due *person-set* determinati da due membri *a* e *b*, tra di loro connessi da continuità psicologica, corrisponde ad una persona solo se non comprendono *person-stages* contemporanei. Se invece i due *person-set* contengono *person-stages* contemporanei allora si considerino tutte le sottoclassi dell'unione ottenute omettendo tutti i membri di ogni gruppo di *person-stages* simultanei tranne uno. Ognuna di queste sottoclassi i cui membri sono collegati da continuità psicologica è una persona.

Secondo questa definizione quindi, in un caso di fissione in cui si vengono a produrre due candidati all'identità con il soggetto di partenza, sarebbero coinvolte tre persone, due delle quali sopravvivono all'operazione, mentre una no. Nel caso invece in cui si abbiano più continuanti con un diverso grado di continuità con l'originale, allora avremmo tante persone coinvolte quanti sono questi continuanti. Attraverso questa definizione Noonan rende conto del fatto che se supponiamo che il referente del termine “Io” sia il *person-stage* che al momento del proferimento della parola è il “marker” del *person-set*, nel momento in cui io contemplo la prospettiva di una fissione, considero una situazione in cui “Io” cesso di esistere, nonostante entrambe le persone risultanti dalla fissione possano veridicamente affermare di essere esistite precedentemente alla fissione.

Noonan cerca in questo modo di rendere conto delle stesse intuizioni alle quali faceva riferimento anche Nozick. In particolare cerca di rendere conto del fatto che plausibilmente dovrei preoccuparmi specialmente solo del mio futuro, mentre dovrei secondariamente preoccuparmi di quello dei miei continuanti in proporzione al loro grado di identità con me. Secondo la visione proposta da Noonan, posto di fronte alla prospettiva di una futura fissione non dovrei preoccuparmi specialmente del futuro dei miei continuanti, ma dovrei piuttosto aver paura per la mia stessa vita. Stando così le cose però, non sembra irrazionale che io paghi un infermiere per distruggere uno dei due prodotti della fissione dopo l'operazione; se la fissione ha luogo infatti, io cesso di esistere, mentre nel caso in cui solo uno dei miei continuanti sopravviva io continuo a vivere come la stessa persona.

Noonan per mostrare che solo apparentemente la sua teoria porta a questa conclusione controintuitiva, considera separatamente le due situazioni possibili. Nel caso in cui io non paghi perché uno dei due prodotti della fissione venga distrutto, io cesso di esistere. Nel caso invece in cui io paghi perché mi si assicuri un solo continuante la situazione non cambia, in quanto a sopravvivere non sarei comunque io, ma solo il continuante risparmiato, l'unica reale differenza tra le due situazioni è che nella seconda sopravvivono meno persone, nulla cambia relativamente alla mia sopravvivenza all'operazione. Anche considerando l'intera situazione dall'altro lato e ipotizzando che io effettivamente abbia corrotto un infermiere per assicurarmi che dopo il processo di fissione solo uno dei miei continuanti sopravviva, allora dopo l'operazione io dovrei identificarmi con l'unico continuante sopravvissuto. Guardandomi alle spalle e pensando alle mie azioni però, quale giustificazione potrei trovare per aver fatto distruggere gli altri continuanti? Secondo Noonan nessuna. Nella situazione possibile scongiurata dalle mie azioni la mia vita non sarebbe stata in alcun

modo più corta; in quella situazione infatti, io sarei stato semplicemente uno dei continuanti sopravvissuti alla fissione e non l'unico. Sembra quindi che Noonan, contrariamente a Nozick, con la sua formulazione sia riuscito a rendere conto delle intuizioni che abbiamo riguardo a queste situazioni, caratterizzando consistentemente il comportamento di chi paga per far distruggere tutti i suoi continuanti eccetto uno come irrazionale.

La tesi di Noonan però non è esente da problemi. Se infatti la sua definizione dell'estensione del termine persona e la sua applicazione di essa ai casi di fissione, sembra avere una qualche plausibilità quando viene osservata da un punto di vista temporalmente successivo al processo, la sua formulazione rivela tutte le sue difficoltà nel momento in cui osserviamo l'operazione da un punto di vista temporalmente precedente ad essa. Quella di Noonan viene infatti chiamata una "multiple occupancy theory" e comporta che nel nostro corpo pre-fissione coesistano almeno tre persone diverse, rivoluzionando la nostra tradizionale concezione che identifica ogni persona come una mente unica ed unificata all'interno di un corpo. La fonte di questa rivoluzione concettuale risiede nel fatto che Noonan, per tentare di associare consistentemente criterialismo psicologico e determinatezza dell'identità, ha adottato un'ontologia quadrimensionalista. Secondo la più tradizionale ontologia tridimensionalista, gli oggetti hanno solo parti spaziali e non temporali, di modo che in ogni momento della sua esistenza ogni oggetto è completamente presente, gli oggetti secondo questa visione ontologica, sono considerati dei *continuant*. In un'ottica quadrimensionalista invece, gli oggetti, e di conseguenza anche le persone, non hanno solo parti spaziali ma anche parti temporali, ovvero ciò che Noonan chiama *person-stages*. Conseguentemente se si adotta un'ontologia di questo tipo, in ogni momento della mia esistenza solo una parte di me è presente, similmente a come solo una parte di me è spazialmente presente nella regione di spazio che si alza dal terreno di soli venti centimetri. Chi difende la validità di un'ontologia di questo tipo quindi, considera gli oggetti come degli *occurrent*.

Se considerata in un'ottica tridimensionalista, la situazione descritta sopra è semplicemente priva di senso. Comporterebbe infatti, che in un tempo precedente alla fissione almeno tre persone completamente presenti occupino la medesima porzione di spazio contemporaneamente, cosa palesemente assurda. Dal punto di vista di un'ontologia quadrimensionalista invece, le tre persone compresenti precedentemente alla fissione condividerebbero semplicemente alcune parti temporali, ma possono consistentemente essere distinte, di modo che l'esistenza dell'una non dipenda dall'esistenza dell'altra. Se infatti all'operazione sopravvivesse solo un emisfero, conseguentemente cesserebbero di esistere due delle tre persone coinvolte nel processo di fissione, senza che questo abbia alcuna ripercussione sull'unica che sopravviverebbe. Rimane comunque molto difficile concepire come tre persone distinte possano condividere una regione spaziale ed un segmento temporale senza che si riscontri nessuna disunità psicologica. Questa difficoltà inoltre appare ancora più evidente se viene comparata con il fatto che il principio secondo il quale ad ogni segmento temporale psicologicamente unificato corrisponde una ed una sola persona, al contrario, mantiene la sua intuitiva validità. Nell'ottica di una simile prospettiva diviene particolarmente problematico rendere conto dell'apparente unicità e coerenza del nostro proferire la parola "Io" in riferimento a pensieri di natura riflessiva corrispondenti ad eventi fisici localizzabili nel cervello pre-fissione. Dal momento in cui si fanno

coesistere più persone all'interno di uno stesso corpo, non si riesce più a spiegare l'unitarietà della coscienza né a rendere conto della intuitiva plausibilità della sua connessione con l'unitarietà delle funzioni del cervello umano.

Dalla formulazione di Noonan emerge anche un altro problema. Sembra infatti, che la coesistenza temporale di due o più persone all'interno di uno stesso corpo, dipenda dal destino di questo. In particolare se il destino di una persona è quello di subire una fissione, allora all'interno del corpo che subirà l'intervento sono compresenti dalla nascita più persone. Se infatti ogni linea di continuità psicologica abbastanza forte rappresenta la storia di una persona allora, dato che il soggetto di partenza ed i prodotti della fissione condividono la parte dei loro ricordi che va dalla loro nascita all'operazione, questi devono aver avuto lo stesso principio di esistenza. Se invece il destino di una persona non è quello di dover subire una fissione nell'arco della sua vita, allora questa mantiene la sua unitarietà. Non solo, secondo l'argomentazione di Noonan, il numero delle persone compresenti all'interno del corpo precedentemente al processo di fissione è determinato dal numero dei continuanti che verranno prodotti dalla fissione. Nel delineare la sua teoria infatti, Noonan dice esplicitamente che, nel caso di un processo di fissione come quello in cui vengono separati i due emisferi di un cervello umano, sono tre le persone coinvolte nell'operazione esistenti anche precedentemente alla stessa. Nel caso invece in cui il soggetto di partenza sia coinvolto in un'operazione che ha come risultato la produzione di un numero maggiore di continuanti, allora le persone coinvolte in questo processo saranno tante quanti sono i continuanti più una. Ciò che credo sia molto difficile da spiegare, anche per chi adotta un'ontologia di tipo quadrimensionalista, è l'implicita predeterminatezza del numero delle persone spazialmente e temporalmente compresenti nel soggetto di partenza precedentemente all'operazione. Sembra infatti che il numero delle persone compresenti nel corpo il cui destino è quello di subire una fissione, sia predeterminato in relazione al tipo di fissione che il corpo è destinato a subire.

Adottare un'ontologia quadrimensionalista comporta inoltre non pochi problemi di per sé. Consideriamo ad esempio il fatto che normalmente attribuiamo alle persone proprietà come: "avrebbe potuto vivere più a lungo", se però le persone sono identiche a delle entità quadridimensionali, ovvero a delle somme mereologiche, queste non possono estendersi temporalmente più o meno a lungo di quanto non facciano. Sembra quindi che le persone, così come gli altri oggetti, posseggano proprietà che delle entità quadridimensionali non sono in grado di possedere. Un altro problema tipico che chi adotta una ontologia quadridimensionale si trova costretto ad affrontare è quello che riguarda la spiegazione della possibilità del cambiamento di un oggetto, ovvero la possibilità di perdere alcune proprietà o di acquisirne di nuove. Secondo la teoria quadrimensionalista in particolare, il cambiamento di un oggetto è un fenomeno interamente riducibile alla successione delle sue parti temporali, a ciascuna delle quali vengono associate determinate proprietà. In quest'ottica si dà un cambiamento nell'oggetto ogniqualvolta ad una sua parte temporale ne succede un'altra con proprietà diverse, se una parte temporale di una persona in un tempo t' possiede la proprietà "essere seduta", mentre una parte temporale successiva della stessa persona in t'' possiede la proprietà "essere in piedi", allora questa persona ha subito un cambiamento. Il problema di questa prospettiva è innanzitutto l'oscurità della nozione di "parte temporale", che non viene mai propriamente definita. Sembra inoltre che i sostenitori di

un'ontologia di questo tipo ammettano conseguentemente la proliferazione di infinite entità, le quali succedendosi l'una all'altra iniziano e cessano di esistere letteralmente dal nulla.

Chi utilizza questo tipo di ontologia infine, nel cercare di rendere conto della possibilità del cambiamento di un oggetto, finisce per annullare la nozione stessa del cambiamento. Nonostante parti temporali diverse possano avere proprietà diverse infatti, ciascuna delle parti temporali di un oggetto inizia e finisce senza che queste cambino, senza che nessuna proprietà venga persa o acquisita. Un sostenitore di questa teoria quindi si trova costretto a fare a meno della nozione di cambiamento, ridotta al solo succedersi di parti temporali diverse, le cose in quest'ottica non sono più in grado di cambiare nel significato tradizionale del termine.

Anche il tentativo di Noonan di sviluppare un criterialismo psicologico che mantenga consistentemente saldo l'*only x and y principle* quindi, non è esente da problemi. Il prezzo pagato per poter superare il *Reduplication Argument* di Bernard Williams inoltre si è rivelato essere molto più alto di quanto possa essere ragionevolmente accettato. Stando così le cose l'obiezione di Williams sembra per il momento mantenere pressoché intatta la sua forza argomentativa che ora si estende esplicitamente fino a riferirsi anche ai criterialismi psicologici fondati sulla continuità del cervello. A mio modo di vedere le considerazioni fatte da Bernard Williams se vengono affiancate ad un'analisi puntuale delle nostre intuizioni riguardo ai casi di fissione della persona, chiudono i sostenitori della sufficienza del criterio psicologico per l'identità personale in un vicolo cieco dal quale è molto difficile districarsi se non a carissimo prezzo. Le conclusioni a cui i criterialisti sono costretti ad arrivare per far fronte alle obiezioni a loro mosse sono infatti fortemente controintuitive e problematiche, sia che si tratti di utilizzare un concetto di identità privo dell'*only x and y principle*, sia che si tratti dell'adottare un'ontologia quadrimensionalista, per non parlare del fatto che spesso, come ad esempio nel caso di Nozick, si arriva a sostenere entrambe le ipotesi.

Nell'ultima parte di questo capitolo verranno prese in considerazione alcune argomentazioni proposte dai non criterialisti al fine di rispondere ad alcune obiezioni che classicamente vengono loro poste. In particolare si cercherà di mostrare che sostenere che non possano esistere criteri necessari e sufficienti per stabilire l'identità personale non comporta l'assunzione di nessuna particolare metafisica o ontologia della persona che la descriva come un ente semplice, indivisibile o addirittura immateriale. Si tenterà infine di mostrare che la difesa di una visione non criterialista dell'identità personale non condanna necessariamente a rendere confuso e ineffabile il concetto di persona, né a rendere priva di significato l'identificazione di una stessa persona come persistente nel tempo.

4.3. Considerazioni finali sul non criterialismo.

Una delle obiezioni che più frequentemente viene mossa contro i sostenitori delle visioni non criterialiste per l'identità personale riguarda il rapporto tra epistemologia e metafisica che queste teorie esprimono. Alle teorie che sostengono l'impossibilità di

analizzare e ridurre l'identità personale attraverso criteri empirici, viene spesso imputato di essere costrette ad accettare una visione della persona intesa come un ente persistente e metafisicamente semplice, un qualcosa di indivisibile, inalterabile e immateriale. Nonostante alcune teorie non criterialiste siano state effettivamente fondate su visioni metafisiche di questo tipo, come ad esempio quella proposta da Richard Swinburne (1984), vi sono ragioni per sostenere che la proposizione di una teoria non criterialista sul piano epistemologico non debba necessariamente implicare una visione metafisica di questo tipo.

Se il non criterismo viene infatti inteso semplicemente come l'idea che la persistenza delle persone nel tempo non possa essere ridotta a criteri empirici più primitivi non c'è motivo per sostenere che questo implichi una particolare visione metafisica e ontologica della persona. Come viene efficacemente fatto presente in Carlo Gabbani (2007) l'idea fondamentale, di natura prettamente epistemologica, alla base delle interpretazioni non criterialiste dell'identità personale è quella che associa alla categoria "identità personale" un referente reale nel mondo, il quale però non può essere ridotto ad altri fatti empirici più primitivi intesi come suoi costituenti designati da altre categorie autonome e più primitive. Esiste quindi un non criterismo la cui tesi è esclusivamente epistemologica ed associabile consistentemente a diverse concezioni metafisiche, che sostiene l'interpretazione secondo cui la categoria "identità personale" detiene lo status di primitiva all'interno dell'apparato concettuale attraverso il quale interpretiamo e descriviamo la nostra esperienza. Questo non significa che questo tipo di visioni non criterialiste siano compatibili con qualsiasi visione metafisica o ontologica della persona, alcune ipotesi metafisiche di tipo riduzionista proposte dai sostenitori di teorie criterialiste o riduzioniste, ad esempio, continueranno ad essere incompatibili con le visioni non criterialiste dell'identità personale, ma rende questa visione epistemologica possibilmente consistente con diverse visioni ontologiche e metafisiche della persona. Se prendiamo ad esempio le teorie non criterialiste proposte da Lowe (2004), Chisholm (1976) o Merriks (1998, 1999), notiamo che nonostante siano tutte classificabili sotto la dicitura "non criterismo" sono fondate su visioni metafisiche profondamente diverse.

Un'altra classica obiezione che viene rivolta contro le visioni non criterialiste è quella che sostiene che rendano ineffabile e confuso sia il concetto stesso di persona sia quello di identità personale. È necessario notare innanzitutto, che per quanto riguarda la formale caratterizzazione del concetto di identità, questa non viene in nessun modo toccata dal rifiuto di accettare l'esistenza di criteri necessari e sufficienti per l'identità personale. Le proprietà che tradizionalmente vengono ascritte alla relazione di identità quali la simmetria, la transitività, l'intrinsecità, la determinatezza eccetera riguardano esclusivamente la natura della relazione di identità. L'assunzione di criteri a necessari e sufficienti perché la relazione di identità tra due enti temporalmente successivi si dia riguarda invece l'esistenza di condizioni in base alle quali l'identità può essere attribuita a due enti di una particolare categoria.

Per quanto riguarda invece la caratterizzazione dell'intensione di "persona" bisogna chiedersi quale sia la relazione tra criteri di ascrizione ad una categoria e criteri di persistenza dell'oggetto appartenente a tale categoria e se sia possibile mantenere i primi rifiutando i secondi. Nonostante in parte sia vero che la presenza di un criterio necessario di ascrizione possa indicare una componente necessaria per la persistenza

dell'ente in questione, è necessario tenere presente che, come viene sostenuto anche da Merricks (1998), essere un membro di una certa categoria e persistere come tale nel tempo, sono caratteristiche diverse di un oggetto. Un criterio di persistenza quindi, non può coincidere con un criterio di ascrizione e non ne è nemmeno una conseguenza necessaria. Lo stesso criterio mnemonico che spesso viene proposto come criterio di persistenza per le persone, non coincide ne segue direttamente da un criterio di ascrizione alla categoria di persone. Rifiutare l'esistenza di criteri empirici di persistenza per le persone quindi, non implica necessariamente che l'intensione del concetto di persona non possa essere efficacemente descritta. Secondo Gabbani il non criterialismo per l'identità diacronica della persona non esclude nemmeno la possibilità di esplicitare i requisiti del tipo logico "persona" in forma modale e di disposizioni. In particolare è possibile anche negando l'esistenza di criteri empirici di persistenza, formulare una caratterizzazione di persona in base alla quale si ascrive a questa categoria non chi possiede effettivamente determinate proprietà e caratteristiche, ma quanti sono semplicemente in grado di averle.

Una efficace caratterizzazione dell'intensione del concetto di persona è comunque possibile anche rifiutando l'esistenza di criteri necessari e sufficienti di ascrizione alla categoria. Sarebbe ugualmente possibile infatti definire efficacemente cosa significhi appartenere ad una determinata classe facendo riferimento ad un insieme di caratteristiche comunemente attribuite agli appartenenti alla categoria, ad esemplari prototipici oppure alla funzione svolta dalla categoria in questione in inferenze ed enunciati universalmente accettati dalla comunità dei parlanti. In questo modo non si intende negare la fondamentale importanza del concetto di persona per la nostra cognizione, ma sottolineare il fatto che affermare che un concetto non abbia criteri necessari di ascrizione non significa affermare che non se ne possano caratterizzare l'intensione o la funzione linguistica. Riprendendo la formulazione di Strawson ad esempio, si può notare che quando sostiene la natura primitiva e basilare per la nostra cognizione del concetto di persona, non intende negare che se ne possa descrivere l'intensione, ma al contrario lo definisce basilare proprio perché possa descrivere efficacemente e riferirsi a delle entità caratterizzate dal fatto che possono a loro essere attribuiti sia predicati fisici sia mentali.

Rimane possibile chiedersi se una prospettiva non criterialista non privi di una giustificazione l'affermazione della persistenza di una persona nel tempo rendendo di fatto ineffabile il concetto di identità diacronica della persona. A questa obiezione però è possibile rispondere mettendo in luce il fatto che sostenere che non possano esistere criteri necessari e sufficienti per l'identità personale non significa sostenere che non possano esistere evidenze epistemiche in grado di giustificare le nostre asserzioni riguardo alla persistenza delle persone nel mondo. Evidenze empiriche e criteri costitutivi sono logicamente indipendenti, è possibile infatti sia che si abbia un criterio costitutivo che non corrisponda ad una evidenza empirica che possa giustificare le nostre affermazioni sulla persistenza di un oggetto, sia che si abbiano evidenze o argomentazioni valide tali da giustificare i nostri enunciati di persistenza, senza che si abbiano a disposizione criteri costitutivi per l'oggetto in questione. A prima vista sembra però piuttosto vero il contrario, ovvero che l'affermare che qualcosa possa avere la funzione di criterio empirico per l'identità personale, presupponga implicitamente un criterio di persistenza. A tale proposito Merricks (1998) riconosce che affermare che

qualcosa possa contare come evidenza in grado di giustificare l'identità personale implica da parte nostra l'avere delle altre conoscenze sull'identità personale, ma nega che queste altre conoscenze debbano necessariamente essere dei criteri necessari. Merricks cita ad esempio il fatto che il nostro considerare le impronte digitali come una evidenza in grado di giustificare un'affermazione di identità personale, implica la nostra conoscenza del fatto che le persone persistono nonostante perdano conoscenza durante il sonno. Questa nostra conoscenza riguardo alla persistenza delle persone però, non può essere considerata un criterio di identità e non ne implica alcuno. Più precisamente, nel momento in cui parliamo di evidenze per l'identità personale, stiamo parlando di fenomeni empirici che nella nostra esperienza sono soliti caratterizzare le manifestazioni o le conseguenze dell'identità personale, senza però poter avere lo status di criteri definitivi per la stessa. Si tratta in poche parole di fatti che, nel nostro mondo possibile e dal punto di vista della nostra esperienza di persone, concepiamo come manifestazioni o conseguenze dell'identità personale ed è in questo senso che non possono avere lo status logico di criteri basilari, necessari e sufficienti; in quanto non sono generalizzabili ad ogni mondo possibile e non giustificano la possibilità di parlare dell'identità personale nei termini di una serie di fatti descritti impersonalmente.

Rimane quindi esclusa al non criterialista, la possibilità di avere una dimostrazione logica dell'identità personale. La ricerca delle evidenze in grado di giustificarla avrà quindi la forma di un'interpretazione di certi fenomeni come manifestazioni o conseguenze dell'identità personale. In questo senso le "evidenze" sono fenomeni per noi concepibili solo in relazione al manifestarsi dell'identità personale. La ricerca di evidenze sufficienti per giustificare l'identità personale però, non permette di avere certezze riguardo al fatto che una persona non persiste, ma solo prove della sua effettiva persistenza. Non potendo in quest'ottica produrre evidenze che giustifichino affermazioni volte a negare la persistenza di una persona, il compito di impedire la possibilità che due persone diverse vengano identificate come la stessa in due tempi diversi viene affidato alle proprietà che tradizionalmente vengono attribuite alla relazione di identità in generale.

Questa brevissima parte conclusiva aveva il solo scopo di mostrare che, diversamente da quanto viene spesso obiettato, rimane aperta la possibilità per un'impostazione non criterialista di formulare una teoria epistemologica coerente, che non implichi necessariamente una particolare concezione metafisica ed ontologica della persona e che riesca consistentemente a definire efficacemente l'intensione dei termini "persona" e "identità personale". Nell'ottavo capitolo di Carlo Gabbani (2004) ad esempio è possibile trovare una teoria formulata su basi concettuali puramente epistemologiche.

Stando a quanto viene affermato nelle parti precedenti di questo testo infine, sembra piuttosto che l'obiezione secondo cui una visione non criterialista condanni i suoi difensori a sostenere una determinata visione ontologica della persona possa in realtà essere rigirata verso chi sostiene che la continuità psicologica sia un criterio necessario e sufficiente per l'identità personale. È stato infatti messo in evidenza che per poter superare consistentemente le obiezioni di Bernard Williams, un difensore di queste teorie sia costretto o a fare a meno dell'*only x and y principle*, oppure ad adottare una ontologia quadrimensionalista.

4.4. Conclusioni.

In quest'ultimo capitolo sono stati presentati, spero il più chiaramente possibile, i problemi che chi cerca di associare consistentemente un criterialismo psicologico con l'intrinsecità della relazione di identità si trova costretto ad affrontare per superare le argomentazioni proposte da Bernard Williams nel suo *Reduplication Argument*. Trovo personalmente che le conclusioni alle quali arriva Noonan per adempiere a questo scopo siano altrettanto controintuitive e problematiche di quelle alle quali arrivano Nozick e Parfit facendo a meno dell'*only x and y principle*. Nonostante infatti negli ultimi decenni le impostazioni che considerano gli enti come delle entità quadridimensionali abbiano trovato molti sostenitori in diversi campi di indagine, ci sono ancora diverse questioni che vengono da queste teorie lasciate aperte. In questo capitolo sono state riportate per ragioni di spazio solamente alcune obiezioni rivolte direttamente alla teoria di Noonan oltre ad alcune argomentazioni a favore della più tradizionale ontologia tridimensionalista che ho ritenuto essere più rilevanti per il problema dell'identità personale. Alla vastità concettuale dei problemi sollevati corrisponde ad ogni modo la vastità della letteratura che è già stata prodotta in merito alla questione più prettamente ontologica ed alla quale rimando per una trattazione del problema decisamente più esaustiva.

CONCLUSIONI.

Se le argomentazioni prodotte in questa tesi sono corrette, allora il *Reduplication Argument* di Bernard Williams si è rivelato ancora una volta stringente e difficilissimo da superare per chiunque proponga una teoria dell'identità personale classificabile come un criterialismo psicologico. Il tentativo operato da Noonan di superare l'argomento relativizzandolo ad un criterialismo psicologico fondato sulla continuità del cervello inoltre, ha avuto a mio modo di vedere l'effetto di estendere esplicitamente anche nei confronti di queste teorie la forza dell'obiezione. Anche un sostenitore di questa sottoclasse del criterialismo psicologico come Noonan infatti, è stato costretto per districarsi dall'argomento di Williams, a sovvertire completamente i tradizionali principi dell'ontologia tradizionale assumendo un ottica quadrimensionista. Come è già stato mostrato infatti, la serie dei casi citati da Williams, anche nella forma rivisitata proposta da Noonan, impedisce ad un difensore della validità del criterio psicologico che considera la relazione di identità come strettamente intrinseca agli oggetti ai quali si applica di trattare i casi di fissione come situazioni in cui uno dei prodotti dell'operazione mantiene l'identità con il soggetto originale. Dal momento che è palesemente assurdo che una persona temporalmente precedente sia identica a due diverse temporalmente successive, l'unica possibilità che rimane aperta quindi è quella di considerare entrambi i continuanti come non identici all'originale. Per poter sostenere che nei casi di fissione la persona originale cessa di esistere, mentre due nuove persone psicologicamente continue con l'originale cominciano la loro esistenza nel momento della fissione, un sostenitore della validità del criterio psicologico fondato sulla continuità del cervello si trova nuovamente costretto ad abbandonare l'*only x and y principle* per formulare una Teoria del miglior continuante. Se decidesse di mantenere valida l'intrinsecità della relazione di identità infatti, si troverebbe nella scomodissima posizione di dover stabilire quanta parte del cervello della persona originale, o comunque quali funzioni del cervello della persona originale al di là della sola continuità psicologica per la quale un solo emisfero è sufficiente, debba sopravvivere perché sopravviva la persona. Tralasciando il fatto che ci troveremmo di fronte ad una situazione simile a quella di dover stabilire esattamente quanti capelli possa avere qualcuno perché si possa considerare "pelato", un difensore di una simile teoria sarebbe costretto ad ammettere che in certi casi patologici come ad esempio gravi forme di ictus in cui una parte sostanziale del cervello muore, la persona che ha avuto l'ictus cessa di esistere ed al suo posto una nuova comincia la sua esistenza.

Volendo scongiurare simili conseguenze Noonan arriva quindi ad asserire che durante un processo di fissione come quello della separazione e del successivo trapianto dei due emisferi di un cervello umano, sono tre le persone coinvolte, una delle quali cessa di esistere al momento dell'operazione, mentre le altre due, che esistevano già

precedentemente all'operazione ed occupavano lo stesso spazio della prima originale, no. Perché questa ipotesi non sia effettivamente assurda come sembrerebbe a chiunque fosse digiuno di filosofia, Noonan è costretto ad adottare un'ontologia quadrimensionalista che alle tre dimensioni spaziali dell'ontologia tradizionale aggiunge una quarta dimensione temporale. Questo gli consente di dividere le persone in parti temporali, similmente a come possono essere divise in parti spaziali, ed a sostenere che le tre persone che secondo la sua teoria condividevano lo stesso corpo prima della fissione hanno condiviso semplicemente alcune parti spaziali per alcune loro parti temporali.

Per riuscire a districarsi dal vicolo cieco in cui Bernard Williams ha chiuso i criterialismi psicologici quindi, Noonan è costretto ad operare una rivoluzione ontologica che, seppure sembra fornirgli una via di uscita dal *Reduplication Argument*, ha delle conseguenze fortemente controintuitive sul nostro stesso concetto di persona che a mio parere sono difficilmente spiegabili. Sembra inoltre, che Noonan pur di salvare la validità del criterio psicologico per l'identità personale mantenendo la coerenza del suo sistema di inferenze, sia stato costretto a riformulare e rimaneggiare l'intero sistema di inferenze attorno al criterio psicologico, con una mossa che ricorda il pensiero di Quine in proposito. Trovo inoltre quantomeno strano che Noonan, nelle sue obiezioni verso le Teorie del miglior candidato, faccia spesso riferimento alle nostre intuizioni del senso comune sul concetto di persona e su quello di identità personale, per poi arrivare a proporre una teoria secondo la quale nello stesso corpo possano convivere più persone che condividono parti temporali e spaziali. A mio parere Noonan per cercare di salvare sia l'intuitiva validità dell'*only x and y principle* ovvero della stretta intrinsecità della relazione di identità sia la validità del criterio psicologico per l'identità personale, è arrivato a proporre una tesi forse ancora più fortemente controintuitiva e problematica.

L'obiettivo di questa tesi era quello di mostrare che chi sostiene che la continuità psicologica sia un criterio sufficiente per l'identità personale, nel momento in cui si trova a dover rendere conto delle nostre intuizioni riguardo ai casi di fissione ed a rispondere alle obiezioni mosse da Bernard Williams, si trova in un vicolo cieco dal quale è in grado di uscire solo al carissimo prezzo dell'intrinsecità della generale relazione di identità o dell'abbandono della tradizionale ontologia tridimensionalista. A mio parere entrambe queste soluzioni sono scongiurabili, sia per quanto riguarda strettamente il problema dell'identità personale, sia per il fatto che hanno conseguenze molto forti e similmente problematiche che si estendono su altri campi di indagine filosofica. Trovo, arrivati a questo punto, che sia molto più sensato semplicemente abbandonare la strada del criterialismo psicologico per tentare altre vie di soluzione al problema.

Nell'ultima parte del quarto capitolo, in particolare ho cercato brevissimamente di evidenziare la plausibilità delle teorie non criterialiste per l'identità personale mostrando come molte delle critiche che vengono comunemente mosse a queste interpretazioni possano in realtà essere agilmente superate. Nonostante il fatto che io evidentemente simpatizzi per questa categoria di visioni del problema ed abbia trovato molto interessante l'interpretazione epistemologica e metafisica del concetto di persona proposta da Strawson (1959), la formulazione di una originale proposta teorica di questo tipo esula dagli scopi e dagli spazi di questa tesi. Concludo semplicemente

sottolineando il fatto che la proposizione di una teoria non criterialista per l'identità personale che sia fondata su basi esclusivamente psicologiche, non escluderebbe una visione ontologica di tipo fisicalista della persona. Come viene effettivamente proposto da Gabbani (2007) nel suo ultimo capitolo infatti, per quanto riguarda l'aspetto estensionale del termine "persona" si afferma che se qualcuno in un dato momento è una persona con un corpo umano, allora se conserva lo stesso organismo umano rimane la stessa persona. Gabbani quindi considera la persistenza dell'organismo fisico umano come una manifestazione adeguata della sua persistenza come la stessa persona. Gabbani nella proposizione della sua originale teoria si astiene dal formulare ipotesi di tipo ontologico o metafisico focalizzandosi sull'aspetto epistemologico della nostra cognizione dell'identità personale. Utilizzando la terminologia di Sellars insomma, Gabbani discute dell'immagine manifesta dell'uomo nel mondo, formulando delle tesi che non entrino direttamente in contrasto con quelle dell'immagine scientifica dell'uomo nel mondo.

BIBLIOGRAFIA.

- Agar, N. (2003), *Functionalism and Personal Identity*, in *Nous* 37:1, pp. 52 – 70.
- Alston, W.P., Bennett, J. (1988), *Locke on People and Substances*, in *Philosophical Review* 97, pp. 25 – 46.
- Aristotele. *Fisica*, Luigi Ruggiu a cura di, Rusconi, Milano, 1995.
- Ayer, A.J. (1963), *The Concept of a Person and Other Essays*, Macmillian, Londra.
- Brown, C. (2006), *Peter Strawson*, Acumen, Stocksfield.
- Chisholm, R.M. (1976), *Person and Object*, Allen & Unwin, Londra.
- Evans, G. (1978), *Vague Objects*, in *Analysis* 38, pp. 208.
- Evans, G. (1982), *The Varieties of Reference*, Clarendon Press, Oxford.
- Gabbani, C. (2007), *Per un'Epistemologia dell'Esperienza Personale*, Guerini e Associati, Milano.
- Garrett, B. (1998), *Personal Identity and Self-Consciousness*, Routledge, Londra.
- Geach, P.T. (1962), *Reference and Generality*, Ithaca, Cornell University Press.
- Leibniz, G.W. (1953), *Discourse on Metaphysics*, trad. P. Lucas and L. Grint, Manchester University Press, Manchester.
- Leibniz, G.W. (1981), *New Essays on Human Understanding*, trad. P. Remnant and J. Bennett, Cambridge University Press, Cambridge.
- Locke, J. (1961), *An Essay Concerning Human Understanding*, ed. J. Yolton, Dent, Londra.
- Lowe, E.J. (2004), *An Introduction to the Philosophy of Mind*, Cambridge University Press, Cambridge.
- Merricks, T. (1998), *There Are No Criteria of Identity Over Time*, in *Nous* 32, pp.106 – 124.
- Merricks, T. (1999), *Endurance, Psychological Continuity, and the Importance*

- of Personal Identity*, in *Philosophy and Phenomenological Research* 59, pp. 983 – 997.
- Noonan, H.W. (1989), *Personal Identity*, Routledge, Londra, (edizione rivisitata 2003).
 - Noonan, H.W. (2012), *personal Identity and its Perplexities*, in *Personal Identity: Complex or Simple?*, ed. G. Gasser e M. Stefan, Cambridge University Press, Cambridge.
 - Nozick, R. (1981), *Philosophical Explanations*, Clarendon Press, Oxford.
 - Olson, E. (1997), *The Human Animal: Personal Identity Without Psychology*, Oxford University Press, Oxford.
 - Parfit, D. (1984), *Reasons and Persons*, Clarendon Press, Oxford.
 - Reid, T. (1941), *Essays on the Intellectual Powers of Man*, ed. A.D. Woozley, Macmillan, Londra.
 - Shoemaker, S. (1963), *Self-Knowledge and Self-Identity*, Cornell University Press, Ithaca.
 - Shoemaker, S. (1970), *Persons and Their Parts*, in *Identity, Cause and Mind*. *Philosophical Essays*, Clarendon Press, Oxford 2003, pp. 19 – 48.
 - Shoemaker, S. (1984), *Personal Identity: a Materialist Account*, in R. Swinburne, S. Shoemaker, *Personal Identity*, Blackwell, Oxford, pp. 67 – 132.
 - Shoemaker, S. (2004), *Functionalism and Personal Identity – a Reply*, in *Nous* 38:3, pp. 525 – 533.
 - Strawson, P. F. (1959), *Individuals. An Essay in Descriptive Metaphysics*, Routledge, Londra.
 - Swinburne, R. (1984), *Personal Identity: the Dualist Theory*, in R. Swinburne, S. Shoemaker, *Personal Identity*, Blackwell, Oxford, pp. 3 – 66, 133 – 138.
 - Wasserman, R. (2012), *Personal Identity, Indeterminacy and Obligation*, in *Personal Identity: Complex or Simple?*, ed. G. Gasser e M. Stefan, Cambridge University Press, Cambridge.
 - Wiggins, D. (2001), *Sameness and Substance Renewed*, Cambridge University Press, Cambridge.
 - Williams, B. (1970), *The Self and the Future*, in *Problems of the Self*.

Philosophical Papers 1956 – 1972, Cambridge University Press, Cambridge 1973, pp. 58 – 78.

- Williams, B. (1973), *Problems of the Self. Philosophical papers 1956 – 1972*, Cambridge University Press, Cambridge.